ASSESSING IMPACTS OF CROP-WILD INTROGRESSION IN LENTIL USING INTERSPECIFIC *LENS* SPECIES RECOMBINANT INBRED LINE POPULATIONS

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of PhD

In the Department of Plant Sciences

University of Saskatchewan

Saskatoon

By

Li-An Chen

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Plant Sciences
51 Campus Drive,
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A8
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

ABSTRACT

Lentil (*Lens culinaris* Medikus) (Lcu) is one of the earliest domesticated plant species. *Lens ervoides* (Ler) is a wild species from the tertiary genepool carrying resistance to multiple lentil diseases. Two interspecific Lcu x Ler recombinant inbred lines (RIL) populations, LR-26 and LR-59, had been developed to help introduce disease resistance to lentil. The hybridization broadened the genetic base and impacted many traits beyond just disease resistance. In this study, I assessed the variability of several important agronomic and seed-quality traits. The goals of this thesis project were to 1) determine the phenotypic variation of traits of agronomic and seed quality importance resulting from introgression; 2) assess the level of introgression of Ler genome based on genetic markers; and 3) perform marker-trait association analysis to identify introgression regions underlying the agronomic and seed quality traits observed.

Seven agronomic traits and five seed quality traits were assessed in multi-environmental field trials across three years. One population, LR-26, was genotyped using genotyping-by-sequencing and the resulting single nucleotide polymorphisms were used to construct a high-density linkage map.

There was a significant genotypic and site-year effect on each trait. A reciprocal translocation involving chromosomes 1 and 5 of Lcu with respect to Ler caused several issues, including marker distortion in the rearranged areas which hinted at a possible selection. Also, heterozygosity that was due to aberrant homoeologous pairing as a result of the translocation and semi-sterility from the presence of the translocation probably made maintenance of population size during RIL development challenging.

Chromosomal rearrangements caused marker pseudolinkage and a really large linkage group (LG1) that corresponds to parts of chromosomes 1, 5 and 7. QTL results showed that the quantitative traits were controlled by multiple minor-effect QTLs which could be used to track the introgression of desirable traits. However, there could

be challenges when selecting for QTLs underlying these rearrangement regions for introgressions using LR-26.

Overall, genome introgression has brought tremendous phenotypic variability and help broaden genetic base of lentil. This study showed the potential and challenges using Ler as a genetic resource for lentil breeding.

Acknowledgements

I would like to first thank my academic supervisor Dr. Kirstin E. Bett for her continuously support and guidance. I really enjoyed the discussions with her and sincerely appreciated her encouragement for me to express my thoughts and questions. I want to thank all my advisor committee members, Dr. Ravi Chibbar, Dr. Mike Nickerson, Dr. Bert Vandenberg for their great feedback and precious suggestions. Thank Dr. Pierre Hucl for serving as my committee chair. I want to thank Dr. Matt Nelson for being the external examiner of my thesis and coming all the way here for my defence.

I want to thank all the member of the pulse crop genetic and breeding group's bioinformatic team, Larissa, Lacey, Carolyn, Yichao and Reynold. Thank you for your great work so I can have the resources for my study. My special thanks to Carolyn for help me handle the GBS data and doing the SNP calling for me. I also want to thank all the lab members, Robert and Akiko, thank you for the assistance and help for my lab work. I thank all the pulse crop genetic and breeding group's field lab members, including Brent and Alison and all the field crew. The hard work couldn't have been done without your help. I want to also thank Dr. Yuan, Crystal and Derek, you all helped me a lot in preparing my reports and presentations.

I want to express my specially thanks to the good friends and peers, Yunfei, Katherine, Lindsay, Vlad, Kendra, Rajib, Kundu, and many more great people I have met in Saskatoon. Words can't express how much I appreciate your encouragement, support and warm friendship. I have learnt so much from you all. I wish one day I could help other as much as you all helped me. Thanks to my families for their support and love, I love you all.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

ANOVA: Analysis of variance

CDC: Crop Development Centre

cM: centiMorgan

CSSF: Crop Science Seed Farm

CWI: Crop-wild introgression

CWR: Crop wild relative

DTE: Days (from sowing) to emergence

DTF: Days (from sowing) to flower

GBS: Genotyping-by-sequencing

GP: Genepool

GxE: Genotype-by-environment

H^2: Broad sense heritability

ICARDA: International Center for Agricultural Research in the Dry Areas

Lcu: *Lens culinaris*

Ler: *Lens ervoides*

LG: Linkage group

LOD: Logarithm of odds ratio

MQM: Multiple QTL mapping

PH: Plant height

QTL: Quantitative trait loci

RCBD: Randomized complete block design

RIL: Recombinant inbred line

RP: Reproductive period

SD: Standard deviation

SNP: Single nucleotide polymorphism

STH: Sutherland experiment farm

TRFO: Total raffinose family oligosaccharides

TSW: Thousand seed weight

CHAPTER 1 INTRODUCTION

1.1 Background of Study

Lentil (*Lens culinaris* Medikus, Lcu) is a diploid (2n=2x=14) self-pollinated species, with a haploid genome size around 4 Gb (Arumuganathan and Earle, 1991). Lentil is cultivated as a cool season pulse crop and it is one of the eight earliest domesticated plant species (Weiss and Zohary, 2011). Most global lentil consumption is in Asia, especially within the Indian subcontinent. Lentil was first introduced into Western Canada in the 1970s and Canada is now the largest lentil producer in the world, followed by India and Turkey (Food and Agriculture Organization of the United Nations, 2016).

Genetic erosion of such an ancient and self-pollinating crop is evolutionarily inevitable under the forces of selection, both artificial and unconscious (Weeden, 2007; Zohary, 2004), or by random effects (Ladizinsky, 1985). To protect against genetic vulnerability under a changing climate and evolution of pathogens, expanding the genetic base through utilization of broad genetic resources is highly desired. Crop wild relatives (CWRs) are the ancestors of modern crop species and are natural sources of genetic variability. Pre-breeding progress using exotic crop genetic resources can be very time-consuming and challenging, thus CWRs are less preferred by breeders (Sharma et al., 2013). The major limitations to utilizing CWRs include linkage drag of traits, either undesirable or of lower-value, which would take many generations of back-crossing to get rid of. This is in addition to sterility and viability issues in the hybrids due to various crossing barriers especially when working with species from different genepools (Harlan and de Wet, 1971). Thus, even though CWR may carry alleles that favor agricultural performance, the value of their introgression is often masked by deleterious alleles.

For decades, lentil breeders have sought novel genetic resources from CWR for critical traits such as disease resistance (Erskine et al., 1994). Previously, a wild species of *Lens* from the tertiary genepool, *L. ervoides* (Ler), was selected for the potential of carrying novel resistance to many globally important lentil diseases, such as anthracnose (*Colletotrichum lentis*) (Tullu et al., 2006), ascochyta blight (*Ascochyta lentis)* (Tullu et al., 2010), stemphylium blight (*Stemphylium botryosum)(*Kant et al., 2017; Podder et al., 2013), and possibly even fusarium wilt (*Fusarium oxysporum f.* sp. *lentis*) (Singh et al., 2017). The pulse crop breeding and genetics group at the University of Saskatchewan has generated a wealth of genetic resources of *Lens* species using embryo rescue systems (Saha et al., 2015). These resources have allowed researchers to introduce desirable traits into the cultivar selection pool (Fiala et al., 2009; Vail and Vandenberg, 2011) to expand the current genetic base (Tullu et al., 2013), and to further exploit the genetic components of crop-wild introgression (CWI).

Previously, it was reported that the resistance to multiple diseases had been successfully introduced into a cultivated background through two Lcu x Ler derived interspecific recombined inbred line (RIL) populations: LR-59 and LR-26 (Fiala et al., 2009; Podder et al., 2012; Vail et al., 2012). Both populations were developed in the Crop Development Centre (CDC) of the University of Saskatchewan for the introgression of desirable disease resistance (Tullu et al., 2013), but clearly other variability would also have been introduced. Although Ler has been used in the current breeding program as a genetic donor of disease resistance, the underlying genetic consequences of the CWI remains unclear and became the goal of this study. In addition, Ladizinsky et al. (1985) observed aberrant recombination in the F_1 of crosses between these two species, and a related study of Zamir and Tadmor (1986) also found distortion in genome segregation within progeny derived from Lcu x Ler crosses. Their studies suggested it was important to further assess the impacts from possible genome rearrangements in the Lcu x Ler derivatives. Thus, this thesis focuses on assessing some of the impacts of CWI

at both the phenotypic and genomic levels using two interspecific RIL populations as examples.

1.2 Hypotheses

1) There are significant genetic effects on the phenotypic variation of traits of agronomic and seed quality importance caused by CWI.

2) The distribution of agronomic and seed quality trait values among interspecific RILs will be normal; and the Lcu and Ler alleles will segregate evenly across the genome.

3) Marker-trait associations can be used to identify Ler introgressions of importance containing genes related to agronomic and seed quality traits.

1.3 Research objectives

The objectives of this study were to:

1) Assess the impact of CWI on specific agronomic and seed quality traits by statistical analysis as well as estimating the genetic heritability of these traits under multiple environmental field conditions using the RIL populations LR-26 and LR-59.

2) Genotype the interspecific population LR-26 using genotype-by-sequencing (GBS) to identify patterns of genome introgression.

3) Use QTL mapping to associate these genotypes with variability in the agronomic and quality phenotypes to assign the sources of variation to the individual parents and identify potential markers for further selection.

1.4 Expected contribution

Through the use of wild genetic resources, breeders may further broaden the selection base of lentil and increase its breeding value. From this study, I set out to assess the impacts and challenges of lentil introgression using Lcu x Ler interspecific RILs and to employ genomic tools for developing a lentil pre-breeding system. The

knowledge gained from this study may be of further help for the study of other *Lens* species when used in interspecific hybridization.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction to lentil

Lentil (*Lens culinaris* Medikus, n=x=7, ~ 4Gb) (Arumuganathan and Earle, 1991) is a cool-season grain legume crop. It is a self-pollinating annual plant, typically with an herbaceous plant type, shallow root system and indeterminate growth habit. Lentil was first introduced to Western Canada in the 1970s as a rotation crop. There has been a rapid increase in the Canadian lentil growing area since the early 1990s which may be attributed to the success in developing better-adapted varieties. Lentils were first evaluated across the entire prairie area of Canada but today more than 90% of Canadian lentil production is based in Saskatchewan. This shift might be due to the disease problems often associated with the higher moisture conditions of Alberta and Manitoba (Carew et al., 2013).

Lentils are classified into three main market classes based on seed size and colour. The first is the red market class, which has orange-red cotyledons and can be further classified by seed size into large, small and extra-small. The green market class is characterized by green seed coats and yellow cotyledons. This type typically has larger seed size compared to the red market class, and can be further divided into large, medium and small seed sizes. The third class is the speciality market class, a group which constitutes a minor proportion of lentil cultivation in the world, including Canada. The group includes French green lentil (marbled seed coat and yellow cotyledons), Spanish brown (gray dotted seed coat and yellow cotyledons) and green cotyledon lentil.

Lentil is grown across three major agro-climates: the temperate prairies typical of Saskatchewan; the sub-tropical savannah typical of South Asia; and the Mediterranean type regions including Australia and around the Mediterranean Sea (Khazaei et al., 2016). Canada and India are the major lentil producers of the world producing a total of

3,233,800 and 1,055,536 tonnes, respectively in 2016, followed by Australia, Turkey and the United States (Food and Agriculture Organization of the United Nations, 2016). Canada is also the largest exporter in the global market while India is largest consumer.

2.1.1 The genetic resources of lentil breeding

The concept of gene pools as defined by Harlan and de Wet (1971) can be used to help categorize related species of crop genetic resources. There are seven species that comprise the *Lens* genus, all with the same number of chromosomes. These species can be classified across four gene pools based on their cross-compatibility and genome similarity (Cubero et al., 2009; Wong et al., 2013; Figure 2.1). The primary genepool, GP-1, includes the domesticated species Lcu and the progenitor species *L. orientalis*. *L. tomentosus* are also included in the primary genetic pool based on genetic similarity to Lcu (van Oss et al., 1997). The species in GP-1 can be easily and successfully crossed with cultivated lentil rather easily to produce fertile progeny. The secondary genepool, GP-2, is comprised of *L. lamottei* and *L. odemensis*. The more-distant, wild species Ler belongs to the tertiary genepool, GP-3. Lastly, *L. nigricans* is placed into the quaternary genepool, GP-4, based on genetic distance as well as the apparent inability to create successful crosses with the cultivated species.

Genebanks are important for providing accessible genetic resources as well as having a role in *ex situ* conservation of genetic diversity. The biggest *Lens* germplasm collection in the world is held by the genetic resources section of the International Center for Agricultural Research in the Dry Areas (ICARDA). They hold more than 10,000 accessions including a small proportion of wild species collections. In addition to this collection, ICARDA conducts lentil improvement for many of the major lentil producers in the developing world (Erskine et al., 2011).

Many studies in lentil have been carried out using a core collection of genetic resources, including wild species, to help identify useful economic traits such as biotic

stress resistance (Fernández-Aparicio et al., 2009; Kant et al., 2017; Laserna-Ruiz et al., 2012; Podder et al., 2013; Tullu et al., 2006; Tullu et al., 2010) or abiotic stress tolerance (Gorim and Vandenberg, 2017); and to investigate variation in phenology and agro-morphology (Erskine et al., 1989; Hoffman et al., 1988; Singh et al., 2014; Tullu et al., 2001; Yuan et al., 2017), and seed composition (Tahir et al., 2012).

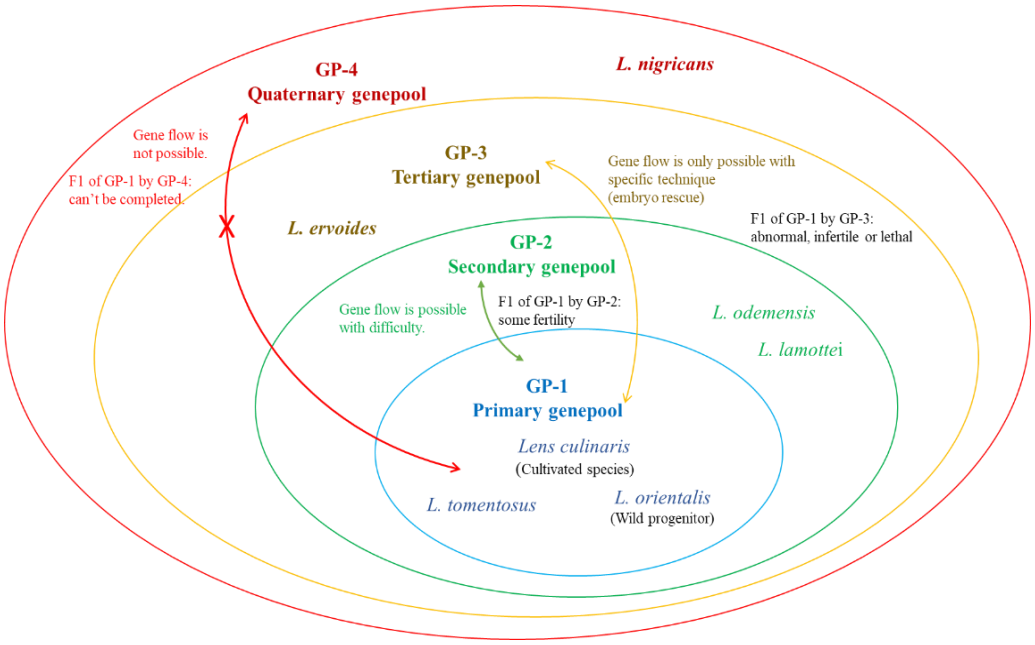


Figure 2.1 Current genepool classification within the genus Lens based on Wong et al. 2013. The arrows indicate possible directions of gene flow to primary genepool. Green arrow indicates gene flow is possible with difficulty. Yellow arrow indicates gene flow is only possible with specific technique such as embryo rescue. Red arrow indicates gene flow is likely not possible.

To make full use of CWI in lentil breeding, in addition to the identification of phenotypic variation, there are other challenges that need to be understood. For instance, overcoming crossing barriers can be a challenge to successful introgression of desirable traits from distantly-related species, and therefore, tissue culture techniques were crucial to the utilization of diversified genetic resources. Cohen et al., (1984) reported the first case of embryo rescue to generate interspecific F_1s of lentil, allowing

more crosses from wider combinations of interspecific populations developed among *Lens* species (Fratini and Ruiz, 2010) to enable research and development.

2.1.2 Lentil seed composition

Like other pulse crops, lentil seeds are especially rich in protein in comparison to cereal grains and vegetable crops. Lentil seeds contain 24 to 30% total protein (Wand and Daun, 2006) and is an important source of vegetable protein for humans. The lentil seed storage protein amino acid composition differs from that of cereal storage proteins in that it has a higher level of lysine and is deficient in sulfur-containing amino acids and tryptophan. Therefore, a meal should include both lentil and cereal grains to ensure complementary sources of vegetable protein (Leterme, 2002). In addition to protein, carbohydrates compose more than 50 % of lentil seed mass, with starch as the major form of carbohydrates. Lentil starch has a higher amylose/amylopectin ratio resulting in a higher level of resistant starch and dietary fibre that compares to those of cereal starches (Joshi et al., 2012). Lastly, lentil seeds contain as low as around 1% of fat. Not only are lentil seeds a rich source of many macronutrients, they are also a significant source of many micronutrients including vitamins such as folate and riboflavin, and trace minerals such as zinc, iron and selenium (Thavarajah et al., 2009; Thavarajah et al., 2011; Kumar et al., 2016). In addition to the nutritive components, lentil seeds also contain several bioactive components, such as trypsin inhibitor, tannins, lectins, saponin and phytate (Muzquiz et al., 2012; Roy et al., 2010; Sharma et al., 1996). These types of secondary metabolites often play a role in response to various environmental stresses (Shalom et al., 1969). However, the consumption of these bioactive components may also lead to certain anti-nutritional effects (Fereidoon, 2014), such as lower protein digestibility (Barampama and Simard, 1993; Sarwar et al., 1989), lower carbohydrate digestibility (Thorne et al., 1983), lower micronutrient bioavailability (Barampama and Simard, 1993) and phytotoxic effects (Podolak et al., 2010). These bioactive components

can often be removed or decreased through various processing methods such as fermentation (Vidal-Valverde et al., 1993), soaking and cooking (Hefnawy 2011; Ruiz et al., 1996), dehulling (Wang et al., 2009) or germination (Ghavidel and Prakash, 2007). Moreover, the bioactive components may provide beneficial health effects as well. For example, polyphenolic compounds provide antioxidant activity with health-promoting benefits (Amarowicz and Pegg, 2008; Oomah et al., 2011).

Like other legumes, raffinose family oligosaccharides (RFOs) can be found abundantly among lentil seeds (Tahir et al., 2011b). These complex carbohydrates exist universally among higher plant species with a higher abundance in legume seeds. Most RFOs are stored in plant seeds, with a lower concentration in plant leaves. In plant physiology, RFOs play a role as protective agents against different types of abiotic stress such as desiccation (Horbowicz and Obendorf, 1994) and cold (Koster and Leopold, 1988). They may also act as carbon sources for seed germination in some species, including lentils (Li et al., 2017; Frias et al., 1996; Vidal-Valverde and Frias, 1992). In human nutrition, RFOs are often considered anti-nutritional because a higher level of consumption could induce a bloating effect due to indigestibility from a lack of α-galactosidase in the mono-gastric gut (Flemings, 1981; Rackis, 1981). However, as was reviewed in Martínez-Villaluenga et al. (2008), a low level of consumption of RFOs may provide health beneficial effect as a source of pre-biotic dietary fibre (Agil et al., 2013; Johnson et al., 2013). The decrease of concentration of RFO in grain legumes may increase the willingness for pulse consumption. Therefore, the selection of lower RFO level can be found in other grain legumes such as soybeans (Dierking and Bilyeu, 2008; Yang et al., 2014; Hitz et al., 2002) and peas (Jones et al., 1999).

2.1.3 Lentil and food and nutrition security

Although the cultivation of lentil occurs in many regions of the world, today most lentil production and consumption is directed towards Asia. This area includes

South Asian countries of India, Nepal and Bangladesh; and West Asian countries of Iran, Syria and Turkey. For lentil growers of developed countries such as Canada, United States and Australia, production is mainly for export. As one of the oldest crop species, lentils play an important role in global food security. Their adaptability to a wide variety of environments as well as their nitrogen-fixation ability makes lentil a low-input and sustainable crop (Suryapani, 2012). The high-protein and low-fat seed profile makes it a staple source of affordable protein in many developing parts of the world, hence often being referred to as "poor man's meat". The starchy lentil seeds are an important source of energy for human consumption, and the straw can be used as animal feed.

Lentil has traditionally been a part of human life in many South Asian countries (Erskine at al., 2011). However, the consumption of lentils and other major pulses have also been increasing in the developed countries, mostly because of the potential health benefits (James and Major, 2002). The consumption of lentils and other dietary pulses may help control obesity through the bulking effect from higher levels of dietary fibre and resistant starch (Kim et al., 2016; McCrory et al., 2010). The low-glycemic index and fat content also make lentils beneficial for the prevention and management of diabetes and a valuable source of energy for diabetic people (Campos-Vega et al., 2010). As well, lentil dietary fibre has prebiotic elements and stimulates the growth and activity of colonic probiotic bacteria (Jounson et al., 2013; Slavin, 2013). Through microbial anaerobic fermentation of undigested carbohydrates, lentil consumption may increase the formation of short-chain fatty acids in promoting anti-inflammatory activity (da S. Queiroz-Monici et al., 2005; Topping and Clifton, 2001). As micronutrient deficiency has become an issue globally (Burchi et al., 2011), the mineral content profile of iron, zinc and selenium of lentil seeds make it a potential source for biofortification (Thavarajah et al., 2009). As pointed out in Podder et al. (2017), lentil has been part of traditional food

sources in many parts of the world, and fortified lentil provides an affordable approach of nutrition level to daily diets.

2.2 Crop genetic improvement

Crop improvement is an activity focused on altering specific traits with a goal of adding value or decreasing the need for management. It is the process of artificial selection to change the direction of evolution by controlling the fitness of progeny of populations. Ever since agricultural activity started, people have selected and kept certain phenotypes that retain higher productivity with easier management and better consumption values. The alteration of the crop genome can be either from the selection of favored or useful traits, or natural selection and environmental adaptation following domestication events.

2.2.1 Plant domestication

The major difference between crop species and their wild relatives is how they rely on human activities to survive and reproduce (Harlan and de Wet, 1975). During the Neolithic era, humans shifted from hunting and gathering to saving and growing wild species which turned into the domestication of crop species. The progress of early agricultural activity arose simultaneously across multiple centers of origin (Harlan, 1971). As a result of domestication, crop species eventually diverged from their progenitor species in both morphological and genetic aspects.

Domestication of crop species is based on selecting and maintaining novel phenotypes to meet human needs (Østerberg et al., 2017). The genes associated with domestication were then fixed within the crop genepool. Domestication syndrome is the outcome of the process; this syndrome varies across various biology backgrounds and organisms. For annual pulse crops, the domestication syndrome usually includes non-dormancy and uniform seed germination, loss of seed-dispersal (pod shattering) mechanisms, larger seed size, loss of anti-herbivore defence mechanisms, and

indeterminate growth habit. These traits are often controlled by a single, or only a few, gene(s) which possibly enabled a rapid domestication (reviewed in Doebley et al. 2006). For example, the basis of shattering resistance in many cereal crop species can be found in an orthologous mutation across the different genomes (Lin et al., 2012).

Domestication followed by seed exchange and migration then spread the cultivated species to diverse environments. Adaptation, selection (conscious and unconscious), founder effects and genetic drift gradually narrowed the genetic bases of crop species

2.2.2 Lentil domestication and adaptation

Lentil is an old-world crop and its domestication can be traced back to the Neolithic era more than 10,000 years ago in the Fertile Crescent near Syria and Turkey (Zohary, 1972). The cultivation of lentil then spread eastward to the Mediterranean regions and became adapted there along with human movements and livestock exchange (Abbo et al., 2005). Lentil then spread from West Asia into the Indo-Gangetic plain around 2,000 BC (Fuller, 2007), where the cultivation of lentil is mainly in the subtropical savannah regions such as northeastern India, lowlands of Nepal and western Bangladesh. The commercial cultivation of lentil was first introduced into North America in northwest USA, in the Palouse region before 1948 (Youngman, 1968), and then further spread to the Northern Great Plains region. Lentil was further introduced to Western Canada prairie in the early 1970s, and Canada is now the biggest lentil producer globally. Another major global lentil producer is southern Australia, where the commercial lentil production started in 1995.

As mentioned earlier, lentils are a cool-season crop and can be grown in a wide range of soil types but prefer well-drained types with low salinity. In the Mediterranean agro-climate, as well as similar climactic zones in Chile and southern Australia, lentils are sown in the fall and harvested as temperatures and day-lengths rise in the spring.

As lentils spread into the Indo-Gangetic region, they adapted to a second agro-climatic zone – the sub-tropical savannah. In this region, lentils are sown after the rainy season and before winter, usually in October, and harvested the following spring. This means that after emergence, plants experience decreasing temperatures and the day-length starts getting shorter until flowering. Then, during the reproductive period, the temperature increases, and the photoperiod gets longer. Lastly, in the temperate areas of the Americas and northern Europe and Asia, lentils are grown as a summer crop and have an opposite growing cycle to that of the South Asian lentils. Therefore, the plants will experience increasing temperature and day-length during the juvenile period and the day-length starts to shorten during the reproductive stage through to harvest.

The start of lentil flowering has been reported to depend on the accumulation of photo-thermal threshold periods (Roberts et al., 1986; Erskine et al., 1990). In corresponding to the various adapted cultivation environments as well as the growing season, the result of dispersal bottleneck also leads to the limited chance of artificial geneflow between materials from different cultivation regions (van de Wouw et al., 2009).

2.2.3 Lentil genetic improvements

Being among the oldest crop species and a staple source of vegetable protein, lentils have been selected by farmers for yield and ease of management for millennia. However, lentils have also remained an underexploited crop species for a long time. Systematic research on lentil improvement was probably not done until ICARDA started a global lentil improvement research program in the late 1970's.

In addition to selecting for high-yielding varieties, the strategy for lentil improvement is affected by their adapted agro-ecological zones. In the winter-sown regions, selection focuses on fit into rain-fed cultivation, with the primary goals being selection for earliness to avoid drought, as well as tolerance for salinity-, heat- and

drought-stress and boron (Erskine et al., 1994). Biotic stress, however, is the major yield-reducer in global lentil production. Fungal diseases of lentils such as ascochyta blight (*Ascochyta lentis*), botrytis gray mold (*Botrytis cinerea*) and fusarium wilt (*Fusarium oxysporum f.* sp. *lentis*), and foliar diseases transmitted by biotrophic pathogens such as powdery mildew and rust (*Uromyces spp*.) are devastating in the global scale (Taylor et al., 2010). Other biotic stresses include parasitic broomrapes (*Orobanche spp.*), parasitic nematode species and pest insects (bruchid beetles and aphids). Beside the abiotic and biotic stresses, selecting for good standability and machine harvest-ability as well as straw yield (for animal feed) are also important goals in many lentil producing countries.

As Erskine et al. (1998) pointed out, the severity of loss of genetic/ allelic diversity in South Asian lentil germplasm is a major limitation to genetic gain in that region. A demand to reconstruct the ancient diversity by using under-utilized genetic resources has become a goal in global lentil improvement projects. The pre-breeding of lentil in ICARDA focuses on screening germplasm collections for many phenotypic variations. Through many base-broadening projects, including wild introgression (Singh et al., 2013), ICARDA as well as local lentil breeders have more recently been releasing improved varieties with a broadened genetic base (Sarker and Erskine, 2006).

The effort in lentil improvement, both genetically and practically, has had many successes globally. According to the statistical database of the United Nations Food and Agriculture Organization, global lentil production has risen from around 850,000 tonnes in the 1961 (the earliest point of record) to nearly 6.3 million tonnes in 2016, with a significant increase in growing area (Food and Agriculture Organization of the United Nations, 2016). Employment of technology to hasten the shift to a genomic-assisted breeding system is important for lentil genetic improvement. Through high-scale molecular mapping and marker development, breeders may develop more efficient breeding goals and achieve more rapid adoption.

2.3 Molecular breeding

Molecular breeding, or genomics-assisted breeding, is based on a combination of plant phenotype and genotype information. It can result in increased efficiency in crop breeding. The techniques can also be applied in estimating crop genetic diversity and further assist conservation of crop genetic resources. Before the availability and improvement in genomic tools, a lack of polymorphic markers was a major limitation to conducting molecular breeding programs in minor crops, including lentil. Today, genome-wide sequencing of a crop species has become more affordable and a very efficient tool for the discovery of high-density molecular markers (Baird et al., 2008; Elshire et al., 2013). In addition, more user-friendly computational methods to deal with the large amount of data points for genome analysis have been developed (Bradbury et al., 2007; Wickland et al., 2017). The genomic era brings scientists a new opportunity to hasten the utilization of underexploited plant species through directional introgression.

2.3.1 Molecular markers for plant breeding

Molecular markers are DNA fragments with polymorphic patterns within populations under selection. DNA polymorphism can be directly linked to specific functional genes to design perfect markers. Alternatively, markers can be used to detect genomic variations without knowing functions. Several types of genomic variations have been used for marker detection assays. Among these, the single nucleotide polymorphism (SNP) is a sequence-based variation and the most abundant type of genomic variation in nature. There are many methods for SNP detection and the more recent technological advance of next-generation sequencing has enabled the detections of large-scale and high-throughput SNP variations across the whole genome. The SNP variations detected from sequencing results can be used to further design PCR-based markers such as cleaved amplified polymorphic sequences (CAPS) and kompetitive allele-specific PCR (KASP) markers, or array-based SNP chips for efficient and cost-

effective genotyping assays. There are many factors to consider when choosing the most suitable marker system in a breeding project. Users need to consider the availability of resources, the cost of assay, the ability to analyze data, the reproducibility of results and the frequency of polymorphism in the genome.

2.3.2 Linkage analysis and quantitative trait loci (QTL) mapping

Quantitative traits are phenotypes that can be measured in a continuous distribution manner, and QTL are the genomic regions associated with these variations. The discovery of QTLs underlying phenotypic variations can help scientists understand the inheritance of, and identify putative genes controlling, complex traits. Identification can be done through QTL mapping, which is the genome-wide scanning of the association of a set of molecular markers with the phenotypic variation found in a bi-parental mapping population. They can also be identified by linkage disequilibrium analysis among unrelated individuals, a technique known as association mapping. Since most agriculturally important phenotypes are controlled by polygenic effects, QTL mapping is very useful in detecting significant markers linked to regions of the genome containing putatively important genes to be used to aid selection of breeding. Sax (1923) first reported the association of seed coat color as a morphologic marker linked with the quantitative variations of seed size in common beans. The first reported QTL mapping method was done using single-marker method by Soller et al. (1976). The marker-linked quantitative effects can be assessed through many different statistical methods such as regression and analysis of variance (ANOVA). Although single-marker method is easy to use and does not required a linkage map in advance, the possibility of recombination between the QTL and a single-marker locus is higher than between the QTL and a flanking marker region. Then the simple-interval mapping method was proposed by Lander and Botstein (1989) by testing the QTL-likelihood ratio using maximum likelihood estimation of all the markers within a putative QTL region.

The disadvantage of simple interval mapping is it only considers one QTL at one time and the result might be biased due to the existence of other linked QTLs on a chromosome using this linear modeling method. To increase the reliability and accuracy of QTL analysis, the use of composite interval mapping aims to detect multiple putative QTL regions by considering all the QTLs in one chromosome separately using a multiple regression model (Jansen, 1993; Zeng, 1993; Zeng, 1994).

2.3.3 Application and limitation of QTL mapping for plant breeding

QTL mapping results are used for cloning candidate genes for functional analysis, or to design specific markers for more efficient selection, namely marker-assisted selection (Lande and Thompson, 1990; Tanksley et al., 1989) that uses the marker-trait association for early stage selection. Furthermore, by carrying on multiple cycles of marker-assisted selection and recombination, breeders may include multiple QTLs into one superior genotype through marker-assisted gene pyramiding (Servin et al., 2004).

The statistical power of QTL mapping could be highly affected by several factors such as the population size, inbreeding generations, marker density and the quality of phenotypic data. These factors include, for example, the labor and resources required to start and maintain a sufficient size of bi-parental population for many generations; as well as the ability to collect the targeted phenotypes under multiple environments trials. It is notable that the only source of genetic variability of a mapping population would come from the parental lines, so it is first required to select useful material and the QTL may not be the same among different populations because of the limited source of variation.

2.3.4 QTL mapping in lentil

Although an ancient crop species, lentil was considered an orphan crop with a lack of genomic tools, until recent advances in molecular technology. Development of

linkage maps provide the source to explore the genome architecture and develop genomic tools for breeders. The first linkage map of lentil was developed in 1984 with only eight allozyme variants and validated with one morphological marker (Zamir and Ladizinsky, 1984).

Due to the lack of genomic information, QTL mapping research on lentil was not conducted until the availability of PCR-based markers. The early studies focused on mapping disease resistance such as ascochyta blight (Ford et al., 1999; Chowdhury et al., 2001), anthracnose (Tullu et al., 2003) and stemphylium blight (Saha et al., 2009), to enable molecular breeding (Ta'ran et al., 2003); while fewer studies focused on mapping agronomic and seed quality QTLs (Fratini et al., 2007; Kahraman et al., 2004). However, with the application of genome-wide, high-resolution linkage analysis using expressed sequence tags (Kaur et al., 2011; Jain et al., 2013; Sharpe et al., 2013), scientists can now establish a high-density linkage map to map complex traits in lentils. As a result, Kaur et al. (2014) used nearly 3000 simple sequence repeat and SNP markers derived from lentil expressed-sequence-tag database to identify a candidate gene for boron tolerance. Fedoruk et al. (2013) used a SNP array to construct a map with seven linkage groups and then mapped multiple QTLs underlying several seed quality traits and flowering time. The application of next generation sequencing technology may hasten the marker-assisted lentil breeding (Kumar et al., 2015). For example, Idrissi et al. (2016) used a combination of random markers, expressed sequence tag-derived simple sequence repeat and SNP markers, as well as genotyping-by-sequencing (GBS) detected SNP markers, to map a total of 18 QTLs underlying several agro-morphologic variations in root and shoot related to drought response. Singh et al. (2017) reported using whole transcriptome sequence analysis to identify putative functional transcripts associated with drought response.

With the availability of the lentil reference genome (variety CDC Redberry) (Bett et al., 2014), analysis of sequencing data has become more accurate and user-friendly.

These advances in technology also allow scientists to explore the wild genomes. For example, Bhadauria et al. (2017) mapped several disease resistance genes within a wild lentil species (*L. ervoides*), which may facilitate marker-assisted introgression breeding in the future.

2.4 Genetic erosion of modern crop species

Crop species have gone through multiple population bottlenecks due to various evolutionary events during the selection pool of modern crop varieties. Genetic erosion may cause vulnerability of plants when facing the threats of pathogen evolution and changing climate. Various threats to genetically vulnerable crop species were documented in the report "Genetic vulnerability of major crops" of the National Research Council, USA (1972). To restore genetic diversity is key to boosting the innate defense system. The utilization of plant genetic resources as the building blocks of breeding programs will allow breeders to bring back genetic and allelic variations lost long ago through multiple bottlenecks.

2.4.1 Pre-breeding: restoring the lost genetic diversity

Broadening the genetic variability of cultivated germplasm is an important goal of crop improvement. Pre-breeders aim to diversify the primary genepool to reduce genetic vulnerability, conserve genetic diversity, and hopefully improve food security. This base broadening approach aims to: select for a specific trait of interest, induce and restore the genetic variation by making wide crosses using various sources, and create introgressions through hybridization (Thormann et al., 2014). Interspecific hybridization can also often lead to transgressive variation (De Vicente and Tanksley, 1993; Rick and Smith, 1953) which will contribute to expand the selection pool. Base broadening through introgression provides breeders with a method of increase the variations of crop improvement (Gur and Zamir, 2004).

Crop wild relatives are the reservoir of ancient genetic diversity. Through pre-breeding, breeders develop germplasm using underexploited genetic resources including CWR (Figure 2.2). Using CWR as a natural genetic resource for introgression breeding has been applied since the last century with many successful examples among the major crop species (Bessey, 1906; reviewed in Dempewolf et al., 2017). A traditional introgression breeding project employs a phenotype-based selection method and genetic donors are chosen so as to contribute useful genes for desirable traits.

While there are a lot of successful examples using the traditional approach, there are other challenges such as the multiple crossing barriers between species that may cause fertility issues, as well as uneven recombination and segregation (Rick, 1969; Zamir and Tadmor, 1986). In addition, linkage drag with undesirable or even lethal effects from unadapted material can slow down the progress of introgression (Brown et al., 1989). The phenotype-oriented approach often underestimates the potential of wild species to bring novel recombination and interactions into the crop genome (Zamir, 2001). A classic example is found in the increase of intensity of fruit color in tomato after

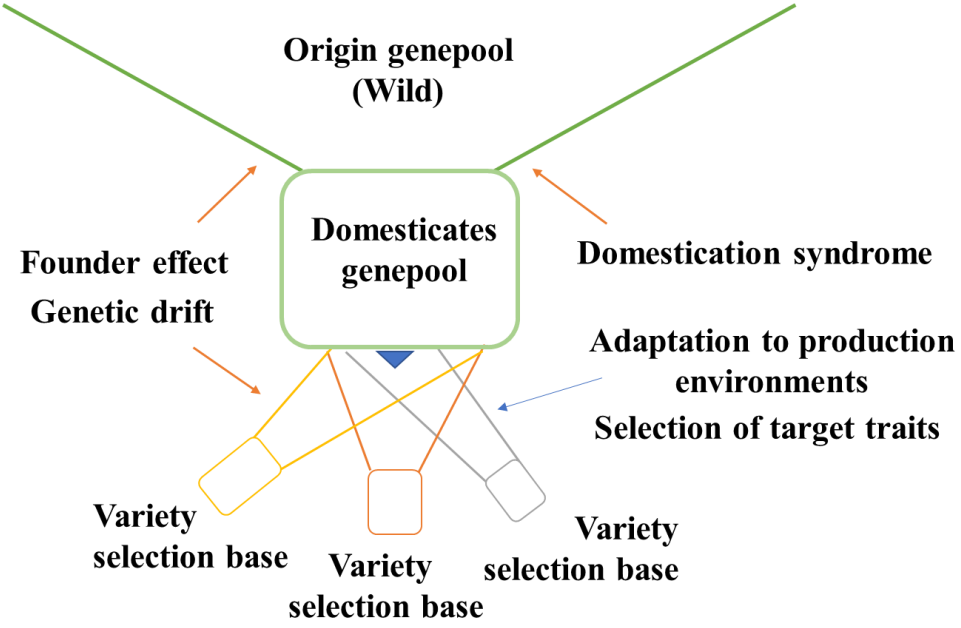


Figure 2.2 The loss of genetic diversity of modern crop varieties from population bottlenecks over time. But the original genetic variability can be restored through pre-breeding.

ripening (Bernacchi et al., 1998). While the wild species didn't have the key enzyme to produce red pigment itself, introgression of a wild gene into the tomato genome led to an increased lycopene level.

According to Dempewolf et al. (2017) and Hajjar and Hodgkin (2007), the major uses of CWRs in crop improvement projects are mostly focused on the introgression of simple traits such as biotic stress resistance and fertility traits such as cytoplasmic-male sterility. The traditional introgression breeding approach can be difficult to apply for complex traits of agricultural importance.

2.4.2 Genomic-assisted introgression

Tanksley and McCouch (1997) emphasized the potential for and application of the "molecular toolbox" for a more efficient and broad utilization of exotic genetic resources. Genetic mapping combined with good phenotyping can help detect significant QTLs or candidate genes from introgression populations, which can then be used to design molecular markers for tracking introgression. There are many applications of genomic-assisted introgression in the improvement of agriculturally important traits such as crop yield (Imai et al., 2013; Xiao et al., 1996), fruit/grain quality (Eshed and Zamir, 1994; McCouch et al., 2007), environmental stress tolerance (Placido et al., 2013) and disease/pest resistance (Zamir et al., 1994) among major crop species. Advancements in genomic and bioinformatic technologies, such as next–generation sequencing are starting to lower the cost to make similar improvements in minor crop species (Varshney et al., 2012).

Development of interspecific populations is a critical step for genomic-assisted introgression breeding. The common approaches are to develop introgression populations such as advanced backcross populations, near-isogenic lines, and chromosome-segment substitution lines, through multiple generations of backcrossing to the recurrent, usually the elite, parent. This is a useful approach aiming to introgress

a single or few genomic segments (Frischa et al., 1998). Another common approach is to create interspecific populations with multiple generations of recombination with many introgression regions. This approach aims to create genome-wide introgression.

As Sharma et al. (2013) pointed out, the use of genebank accessions in crop breeding remains very low. One of the reasons for this low use is the lack of in situ characterization and evaluation of the germplasm. Phenotypic assessment data collection of both CWRs and the interspecific progeny in a cultivated environment as well as information sharing, is believed to be the next major task for more precise introgression-by-design breeding (Dempewolf et al., 2017; Thormann et al., 2014; Zamir 2013).

2.4.3 Background of LR-59 and LR-26

The importance of preservation of wild germplasm for lentil breeding was first proposed by Barulina in 1930. The continuous improvement of Canadian lentils at the Crop Development Centre, University of Saskatchewan, has resulted in expansion of lentil supply globally, while the intensified breeding also resulted in a narrowing of the selection pool. A series of pre-breeding projects were initiated with the aim of using CWR as a source of disease resistance. A series of surveys of core and sub-core collections done to identify novel resistance sources to various diseases such as ascochyta blight (Tullu et al., 2010) and stemphylium blight (Podder et al., 2013), revealed Ler as a high-value carrier of resistance to multiple lentil diseases including anthracnose (Tullu et al., 2006).

2.4.3.1 LR-59: Anthracnose, caused by *Colletotrichum lentis* race 0 and race 1 in the Canadian prairies, is a major disease and yield-reducer of lentils (Buchwaldt et al., 2004). Due to the lack of sources of resistance to anthracnose race 0 among the CDC breeding material, Tullu et al. (2006) developed a core collection including all taxa of *Lens* spp. with materials received from ICARDA, United States Department of

Agriculture-Agricultural Research Service (USDA-ARS) and the Rehovot Institute in Israel. After inoculation of 574 accessions in greenhouse and 484 accessions in a field trial from the sub-core collection with a mixture of race 0 and race 1, a wild lentil accession, *L. ervoides* (Ler) line L01-827a, showed the highest level of resistance to both races of the pathogen. To transfer the resistance to cultivated lentils, a bi-parental, interspecific RIL population, LR-59, was developed from cross between a disease susceptible *L. culinaris* (Lcu) variety (Eston) and this Ler accession (Fiala et al., 2009). To overcome the crossing barrier with a tertiary genepool parent, both ovule and embryo rescue were performed, and 150 plants at the F_2 generation were advanced using single-seed descent for population development. The authors reported segregation for disease resistance among LR-59 RILs as well as some segregation distortion. A reduction in population size from the F_2 to F_7 generations, likely due to a partial sterility issue, was observed in this study resulting in a total of only 85 stable RILs.

2.4.3.2 LR-26: To further explore the potential of wild introgression of Lcu x Ler, another bi-parental, interspecific RIL population, LR-26, was produced. This bi-parental, RIL population was developed using another anthracnose-resistant Ler accession, IG 72815, crossed with the most cross-compatible Lcu variety, Eston (Tullu et al., 2013). After ovule and embryo rescues, 20 F_1 cuttings were grown. Around 300 F_2 seeds were advanced to the F_7 generation using single-seed descent. The LR-26 population consists of a total of 185 RILs that had been bulked at the F_8 generation.

Tullu et al. (2013) reported, from a greenhouse-based trial of Eston and IG 72815, various important agro-morphological traits such as flowering time, plant height at maturity, pod size and seed weight. The contrast between parents was revealed as segregating phenotypes within the RILs. The results implied that, through the future development of a molecular map of LR-26, the successful hybridization of introgressions underlying many complex phenotypic variations could be tracked to multiple genome segments. Base broadening through such a distant-cross created

different level of transgressive segregations in various traits including anthracnose resistance. The result has indicated the potential to harness breeding diversity in lentil.

2.4.4 Summary

For phenotypic characterization, field-based trials of the interspecific populations LR-26 and LR-59 were needed to better understand the impact of CWI. At the genomic level, two issues have been reported among the Lcu x Ler populations: a) non-homologous pairing between Lcu and Ler (Ladizinsky et al., 1985) from a possible chromosomal translocation (Bhadauria et al., 2017; Gujaria-Verma et al., 2014); and b) allelic segregation distortions (Zamir and Tadmor, 1986). Therefore, the need for genetic analysis to help explore the interspecific genomes as well as to develop molecular tools for marker-assisted introgression is of utmost importance.

CHAPTER 3 FIELD-BASED PHENOTYPIC CHARACTERIZATION OF SEVERAL IMPORTANT AGRONOMIC TRAITS IN TWO *Lens culinaris* x *L. ervoides* RECOMBINANT INBRED LINE POPULATIONS

3.1 Introduction

Population bottlenecks imposed by domestication, adaptation and selection have narrowed the genetic base of cultivated lentil (*Lens culinaris*, Lcu) and limited the genetic variation available for improvement (Erskine et al, 1998). CWRs are a natural genetic reservoir and potential genetic resource. CWI is valuable for crop genetic improvement for introducing desirable traits. Despite the lower utility of some agricultural traits in wild genotypes, genomic introgression provides opportunities for novel genetic combinations and recombinations, which can expand the diversity of the primary genepool.

The tertiary genepool species *L. ervoides* (Ler) was shown to carry desirable resistance to several diseases, such as anthracnose (Tullu et al., 2006), ascochyta blight (Tullu et al., 2010), and stemphylium blight (Podder et al., 2012), that were not found in the primary genepool. For anthracnose resistance, two interspecific recombinant inbred line (RIL) populations, LR-26 and LR-59, were developed at the CDC, University of Saskatchewan (Fiala et al., 2009; Tullu et al., 2013) and some selections for improved disease resistance have already been made from these introgression lines.

To further assess the impacts of CWI beyond disease resistance, the objective of the work presented here was to investigate the segregation of agronomically important traits in these two interspecific RIL populations. In this chapter, data from several traits of agronomic importance are presented, including days to emergence (DTE), days to flower (DTF), vegetative period (VP), reproductive period (RP) and plant height (PH) from both LR-26 and LR-59 grown in multiple environment field trials. The results were used to assess the variability of the RILs under field conditions in Saskatchewan and to

estimate broad-sense heritability. Phenotypic data collected from one population (LR-26) was further combined with genotypic data in a marker-trait association study reported in Chapter 6.

3.2 Materials and methods

3.2.1 Plant material

Two interspecific Lcu x Ler RIL populations, LR-26 and LR-59, were used in the study. Eston, a Canadian lentil variety and the market standard for small green lentils, was used as the maternal parent in both crosses (Slinkard, 1981). The paternal parents of LR-26 and LR-59 were Ler accessions IG 72815 and L01-827A, respectively. IG 72815 is from Turkey and was received from the seedbank collection of the International Center for Agricultural Research in Dry Areas (ICARDA) (Tullu et al., 2006). L01-827A was derived from a single Ler plant selected as a variant in *L. orientalis* accession PI 72847 based on its resistance to anthracnose race 0 (Fiala et al., 2009). The F_1 hybrids of both interspecific combinations were produced through ovary and embryo rescue to break the crossing barrier as described in Fiala et al., (2009) and Tullu et al. (2013). Each RIL population was established using single seed descent thereafter until the F_7 generation when the seeds were bulked from single plants and F_7- derived RILs with at least three additional generations of selfing were made available for this study, resulting in 67 RILs of LR-59 and 172 RILs of LR-26.

All trials were set up as a three-replicate randomized complete block design (RCBD). Field randomization was done using Agrobase Generation II® (Agronomix Software Inc.). Each plot represented one replication of the genotype. Twenty seeds of were used for each plot, and all seeds except Eston (Lcu) were scarified before sowing to break any potential physical dormancy due to seed coat impermeability. Scarification of seeds was done manually with razor blades for CSSF13 and with a wood polisher for CSSF14, CSSF15 and STH15 (Figure 3.1).

The field trials successfully established in four environments (Table 3.1). The first site was at the University of Saskatchewan Crop Science Seed Farm (CSSF) seed farm in Saskatoon (52°08'N 106°37' W). Trials were conducted there in 2013, 2014 and 2015, and are labeled CSSF13, CSSF14 and CSSF15. The second site was located approximately 12 km northeast of the CSSF seed farm at the Sutherland experiment farm (STH) (52°10'N 106°30'W). The STH trial was conducted in 2015 and labeled as STH15 (Table 3.1).

For CSSF13, no pre-treatment of seeds was made except for scarification (Figure 3.1). To improve the seed germination in subsequent trials, all the seeds were stored at -20℃ for two days following scarification before seeding at CSSF14; and for 7 days at 4 ℃ and 80% humidity for CSSF15 and STH15. The CSSF13 trial was seeded on May 17, 2013. For CSSF14, the trial was seeded on May 22, 2014, but due to a mechanical issue, the second tray (Tray 2, only containing seed from LR-26) was re-seeded on May 23. A trial was originally planted at STH in 2014 but was abandoned due to unclarified but severe disease/damage. In 2015, the seeding dates at CSSF15 and STH15 were May 8 and May 22, respectively.

Table 3.1 Site-years of field trial environments for phenotyping of LR-26 and LR-59.

Environment	Site	Year	Planting date
CSSF13	Crop science seed farm (CSSF)	2013	2013-05-17
CSSF14	Crop science seed farm (CSSF)	2014	2014-05-22, tray 2 seeded on 2014-05-23
CSSF15	Crop science seed farm (CSSF)	2015	2015-05-08
STH15	Sutherland experiment farm (STH)	2015	2015-05-22

Field trials were set up in plots that were grouped as four rows (A, B, C, and D), each representing one hill. At CSSF13 and CSSF14, hills were grown in all 4 four rows (Figure. 3.2 I) while as at CSSF15 and STH15, hills were grown at only the A and C rows (Figure. 3.2 II). There were approximately 30 cm between each set of hills and 30 cm between each row. All seeds were sown approximately 3.8 cm deep. Parents were included in all replications. Plants of the same genotype in each plot were bagged with white mesh bags at the full bloom stage to minimize seed loss due to dehiscence (Figure 3.2 III). At CSSF15 and STH15, ropes were strung above the rows and the mesh bags were tied on using clips to provide better support (Figure 3.2 IV). Standard lentil plot

Figure 3.1 Seed scarification.
Left (I): Scarified seeds. Right (II): A wood polisher was used for scarification.

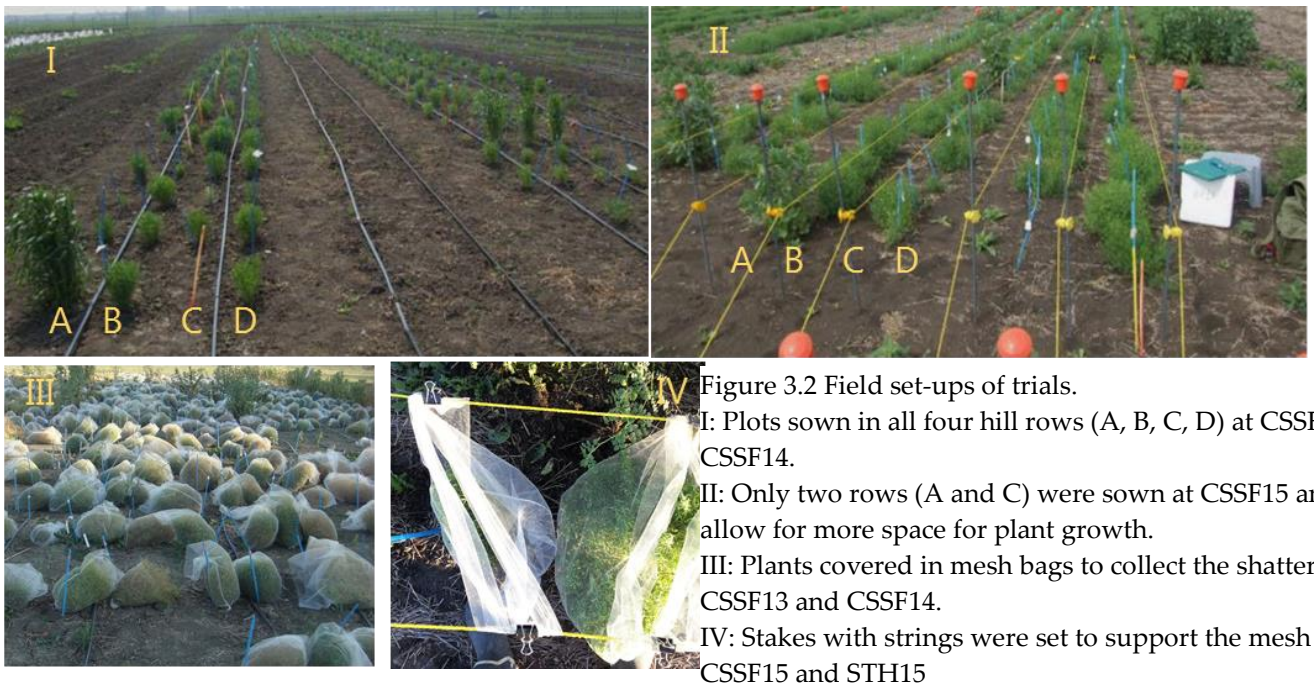


Figure 3.2 Field set-ups of trials.
I: Plots sown in all four hill rows (A, B, C, D) at CSSF13 and CSSF14.
II: Only two rows (A and C) were sown at CSSF15 and STH15 to allow for more space for plant growth.
III: Plants covered in mesh bags to collect the shattered seeds at CSSF13 and CSSF14.
IV: Stakes with strings were set to support the mesh bags at CSSF15 and STH15

management methods of the CDC Pulse Crop crew were followed. Pesticide usage was according to the situation determined by the occurrence of invasive pests or disease. The in-season rainfall and temperature at the tested site/year can be found in Appendix A.

3.2.3 Characterization of phenotypic data

Five segregating, quantitative agronomic traits and two segregating, morphological traits were evaluated in the field trials across the four environments. Evaluation was done based on the average performance of a genotype per hill as described below.

3.2.3.1 Quantitative traits

Days to emergence (DTE): The period, in days, from sowing to 50% emergence (evaluated for CSSF13, CSSF14, CSSF15 and STH15).

Days to flower (DTF): The period, in days, from sowing till the majority of plants within a hill were in full bloom (evaluated for CSSF13, CSSF14, CSSF15 and STH15).

Vegetative period (VP): The period between DTE and DTF. This trait was calculated in days for CSSF13, CSSF14, CSSF15 and STH15.

Days to Maturity (DTM): The period, in days, from sowing until 80% of the fully developed pods within a hill reached maturity. Proper stage of rating maturity was missed at CSSF13 so this trait was only recorded for CSSF14, CSSF15 and STH15.

Reproductive period (RP): The period between DTF and DTM. This trait was calculated in days for CSSF14, CSSF15 and STH15.

Plant height (PH): The average length, in cm, of the plant (measured from the first node to the tip of the main shoot during pod development) of two to three randomly chosen plants in each plot taken at flowering. This trait was measured for CSSF13, CSSF14, CSSF15 and STH15.

3.2.3.2 Qualitative traits

Flower colour: The wing petal color segregated in LR-26 and LR-59 into two types. The 'non-purple' flowers were those with white wing petals bearing blue veins (as found in the Lcu parent); and the 'purple' flowers had purple to light purple wings (as found in the Ler parents). Scoring was done at the full bloom stage at CSSF14 and confirmed in CSSF15 and STH15.

Pod dehiscence: Pod dehiscence was first scored at maturity in the field in CSSF13 and CSSF14, but due to the technical difficulties, this trait was eventually scored from the threshed pods in CSSF15 and STH15. Pod dehiscence was scored as resistant or dehiscent based on pod twisting/spiral coiling shape. The plants with majority of the pods split and the valves shaped into twisting coils after maturation and dry down (Figure 3.3, I) were rated as dehiscent type. Plants were rated as resistant when pods remained attached, or some pods split but did not turn into spiral coils (Figure 3.3, II).

Figure 3.3 Pod dehiscence in lentil. Left (I): Dehiscence pods. Pods shattered with twisting/spiral coiling shape. Right (II): Dehiscence resistant pods. After fully matured and dry, some pods may split but the split valves did not form spiral coils.

3.2.4 Data analysis

Box and whisker plots were used to represent frequency distributions. Plots were generated using the software R version 3.3.0 (R Development Core Team) based on an average value of three replications per genotype. For the quantitative traits of both LR-26 and LR-59, the statistical analysis for the analysis of variance (ANOVA) was conducted using SAS software version 9.4 (SAS Institute Inc., USA). To test the ANOVA

assumptions of normality and homogeneity, tests were done for each environment. Normal distribution was tested using the Shapiro-Wilk test of least-squared means of residual distribution SAS Proc Univariate function (Shapiro and Wilk, 1965). A Levene's test was done using the Proc GLM function to test the homogeneity of variance among the samples (Levene, 1960).

The SAS Proc Mixed function was used for all ANOVA tests. The effects were tested for each environment and across multi-environments to estimate genotype-by-environment (GxE) interaction. Genotype was always treated as a fixed factor. For each single environment, the block effect (considered as the effect of replication) was treated as random. For the combined multi-environment test, the blocks were nested within the environments. The blocks (nested within environment), environments and GxE were treated as random factors.

The SAS Proc Varcomp function was used to estimate variance components and to determine the broad sense heritability (H^2) of each trait. H^2 was estimated using the phenotypic values of each plot of RILs from both LR-26 and LR-59 and calculated using Equation 3.1 where $\sigma^2 g$ indicates the genotypic variance, σ^2_{ge} indicates variance of the interaction between the general environment and specific genotypes and $\sigma^2 \varepsilon$ is the variance from the residuals. The symbols "e" and "r" represent the number of environments and the number of replications, respectively. The broad sense heritability was based on the effect of total genetic variance, which included additive, dominance, and epistatic effects.

$$H^2 = \frac{\sigma^2 g}{\sigma^2 g + \sigma^2 ge/e + \sigma^2 \varepsilon / er} \quad \text{Equation 3.1}$$

For the qualitative traits segregating in both LR-26 and LR-59, chi-square tests were used to determine if the segregation fit the expected Mendelian segregation ratio

of 1:1 for single gene control using the SAS Proc Freq function. A two-gene control model was also tested for the segregation of pod dehiscence with the expected segregation ratio of 3:1.

3.3 Results

3.3.1 Variability in quantitative agronomic traits among LR-26 and LR-59 RILs

Tables 3.2 and 3.3 summarize the results of the ANOVA for agronomically important quantitative traits from the multi-environment trials of LR-26 and LR-59, respectively. These traits included the phenological characteristics DTE, DTF, VP and RP, and also the plant structural characteristic, PH. While the genotypic effect of DTE, DTF, VP and PH were all highly significant ($p<0.001$), the genotypic effect of RP was not significant ($p = 0.35$ in LR-26, $p = 0.31$ in LR-59). However, a highly significant GxE interaction was observed for all five traits in both populations, signaling the need to analyze the variation at each environment separately.

The F-test results of the ANOVA from each environment for LR-26 and LR-59 are summarized in Tables 3.4 and 3.5, respectively. DTE and DTF are both important phenological characteristics, and a highly significant genotypic effect ($p< 0.001$) for each of these two traits was observed in all environments, except CSSF15. In both LR-26 and LR-59, the genotypic effect was not significant ($p= 0.13$ in LR-26; $p= 0.15$ in LR-59) for DTE in CSSF15; the genotypic effect for DTF was slightly significant ($p <0.05$) in LR-26 and significant ($p <0.01$) in LR-59 (Tables 3.4 and 3.5). This difference in DTE and DTF observed in CSSF15 may be attributed to a cutworm infestation which appeared approximately two weeks after sowing. An insecticide application allowed for the emergence time of secondary shoot growth of the later emerging plots to be taken. Due to the artificially late DTE results for this specific environment, and given that the ANOVA test of CSSF15 revealed a distinct impact due to the disturbance (Tables 3.4 and 3.5), the DTE data of CSSF15 were removed from the multi-environment ANOVA

tests (Tables 3.2 and 3.3). DTF showed a slightly significant ($p<0.05$) genotypic effect at CSSF15. In both LR-26 and LR-59, a highly significant ($p<0.001$) genotypic effect for VP was observed in all tested environments (Tables 3.3 and 3.4).

The RP is the difference between DTF and DTM and characterizes the required period for maturation. Among the interspecific RILs, the RP values were less consistent across environments. From the F-test result of ANOVA for RP (Tables 3.4 and 3.5), the genotypic effect in LR-26 did not have a significant impact in CSSF14 ($p = 0.09$) and was only slightly significant for LR-59 ($p <0.05$) at that location. But in CSSF15, the genotypic effect was significant ($p<0.01$) for LR-26 and highly significant ($p <0.001$) for LR-59. At STH15, the genotypic effect was highly significant ($p<0.001$) for both populations. Another important agronomic trait segregating among LR-26 and LR-59 RILs was PH. In both LR-26 and LR-59, PH had a highly significant ($p<0.001$) genotypic effect at each of the tested environments (Tables 3.4 and 3.5). A significant replicate/ block effect was observed among all tested traits in the larger population, LR-26, but not in the smaller population, LR-59 (Tables 3.4 and 3.5).

Table 3.2 F-test results from an ANOVA of five agronomic traits recorded for LR-26 RILs grown in four environments in Saskatchewan, Canada.

	Days to emergence [J]		Days to flower		Vegetative period		Reproductive period		Plant height	
Effect	df	F Value	df	F Value	df	F Value	df	F value	df	F value
Genotype	173	4.32***	173	6.00***	173	2.54***	173	1.0ns	173	5.95***
Environment	2	6.97ns	3	38.88***	3	2.85*	2	379.93***	3	31.68**
Block|Environment	6	16.29***	8	12.25***	8	1.30ns	6	1.33ns	8	18.58***
Genotype*Environment	346	1.62***	511	1.49***	511	2.17***	338	1.46***	513	2.07***
C.V. (%)	29.46		12.71		15.09		28.90		30.19	

***: significant at P< 0.001; **: significant at P< 0.01; *: significant at P< 0.05; ns: not significant; df: degree of freedom; C.V.: coefficient of variation; J: DTE from CSSF15 was removed from the analysis

Table 3.3 F-test results from an ANOVA of five agronomic traits recorded for LR-59 RILs grown in four environments in Saskatchewan, Canada.

	Days to emergence [J]		Days to flower		Vegetative period		Reproductive period		Plant height	
Effect	df	F Value	df	F Value	df	F Value	df	F value	df	F value
Genotype	67	1.99***	67	3.17***	67	3.14***	67	1.11ns	67	12.98***
Environment	2	2.32*	3	43.38***	3	16.88***	2	9.01***	3	17.77***
Block|Environment	6	5.04*	8	5.12***	8	1.12ns	6	0.79ns	8	7.09***
Genotype*Environment	134	1.79***	201	1.93***	201	2.18***	132	1.83***	201	1.66***
C.V. (%)	23.82		14.35		13.37		24.24		34.83	

***: significant at P< 0.001; **: significant at P< 0.01; *: significant at P< 0.05; ns: not significant; df: degree of freedom; C.V.: coefficient of variation; J: DTE from CSSF15 was removed from the analysis

Table 3.4 F-test results from an ANOVA of the quantitative agronomic traits recorded for LR-26 RILs at each of four growing environments.

Environment	Source	Days to emergence		Days to flower		Vegetative period		Reproductive period		Plant height	
		df	F Value	df	F Value	df	F Value	df	F value	df	F value
CSSF13[í]	Genotype	173	2.27***	173	4.85***	173	3.4***	--	--	173	13.85***
	Block	2	15.69***	2	4.61ns	2	0.25ns	--	--	2	4.83ns
CSSF14	Genotype	173	3.85***	173	6.12***	172	2.69***	172	1.19ns	172	5.46***
	Block	2	18.2.2***	2	1.87ns	2	11.13***	2	0.91ns	2	15.30***
CSSF15[£]	Genotype	--	--	168	1.36*	168	2.20***	168	3.13**	168	2.14***
	Block	--	--	2	14.68***	2	0.29ns	2	8.03***	2	14.38***
STH15	Genotype	173	3.05***	172	7.82***	172	4.84***	171	5.91***	173	2.61***
	Block	2	12.40***	2	6.12**	2	3.37ns	2	3.49ns	2	32.12***

***: significant at P< 0.001; **: significant at P< 0.01; *: significant at P< 0.05; ns: not significant; df: degree of freedom; CSSF13: Crop Science Seed Farm at 2013; CSSF14: Crop Science Seed Farm at 2014; CSSF15: Crop Science Seed Farm at 2015; STH15: Sutherland experiment farm at 2015; í: RP from CSSF14 was not characterized in this study; £: DTE from CSSF15 was removed from the analysis.

Table 3.5 F-test results from an ANOVA of the quantitative agronomic traits recorded for LR-59 RILs at each of four growing environments.

		Days to emergence		Days to flower		Vegetative period		Reproductive period		Plant height	
Environment	Source	df	F Value	df	F Value	df	F Value	df	F value	df	F value
CSSF13 í	Genotype	67	1.02^{ns}	67	3.04^{***}	67	2.8^{***}	--	--	67	8.99^{***}
	Block	2	0.07^{ns}	2	0.76^{ns}	2	1.2^{ns}	--	--	2	2.05^{ns}
CSSF14	Genotype	67	3.67^{***}	67	4.60^{***}	67	5.18^{***}	67	1.62^{*}	67	7.90^{***}
	Block	2	2.01^{ns}	2	1.97^{ns}	2	0.02^{ns}	2	0.14^{ns}	2	7.07**
CSSF15£	Genotype	--	--	67	1.97^{**}	67	2.30^{***}	67	4.08^{***}	67	3.61^{***}
	Block	--	--	2	5.95^{**}	2	1.17^{ns}	2	1.91^{ns}	2	11.79^{**}
STH15	Genotype	67	1.88^{***}	67	3.44^{***}	67	4.67^{***}	65	2.29^{***}	67	6.46^{***}
	Block	2	4.51^{**}	2	1.14	2	1.49^{ns}	2	1.08^{ns}	2	0.04^{ns}

***: significant at P< 0.001; **: significant at P< 0.01; *: significant at P< 0.05; ns: not significant; df: degree of freedom; CSSF13: Crop Science Seed Farm at 2013; CSSF14: Crop Science Seed Farm at 2014; CSSF15: Crop Science Seed Farm at 2015; STH15: Sutherland experiment farm at 2015; í: RP from CSSF14 was not characterized in this study; £: DTE from CSSF15 was removed from the analysis.

3.3.2 Frequency distributions and variation in agronomic traits among LR-26 and LR-59 RILs across all environments

Quantitative segregation was observed among the RILs of both populations for all five agronomic traits measured (Figures 3.4 - 3.13). The parents of LR-26 had differential performance for all traits at all environments except for the VP of CSSF13 and CSSF14. The parents of LR-59 showed contrasting performance for all quantitative traits measured except VP, where the parents showed similar performance in all environments.

In LR-26, the frequency distributions of DTE across all environments showed skewness towards earlier emergence (Figure 3.4). In the smaller population, LR-59, DTE was even more skewed towards earliness (Figure 3.5). In STH15, all the RILs had an earlier emergence date than the Ler parent. In both LR-26 and LR-59 (Figures 3.4 and 3.5), the emergence was delayed in the CSSF15 tests. DTF had a smaller range of variation among the RILs compared to other phenological characteristics. The VP of the Lcu and Ler parents were similar at CSSF 13 and CSSF14 in LR-26; and at CSSF13, CSSF 15 and STH15 in LR-59. However, the RILs within both populations segregated for VP with transgressive variation mostly towards a longer period (Figures 3.8 and 3.9). Overall, in both LR-26 and LR-59, transgressive segregants could be found for both DTF and VP.

Due to the indeterminate growth habit, the appropriate stage to start tracking maturation time was missed in CSSF13 for both populations. However, for the subsequent years, RP was calculated. From the frequency distribution plots (Figures 3.10 and 3.11), RP was more delayed in both populations at CSSF14 than at CSSF15 and STH15. Plant height (PH) in both populations showed high levels of variability among the RILs and the distribution revealed transgressive segregation in both populations (Figures 3.12 and 3.13).

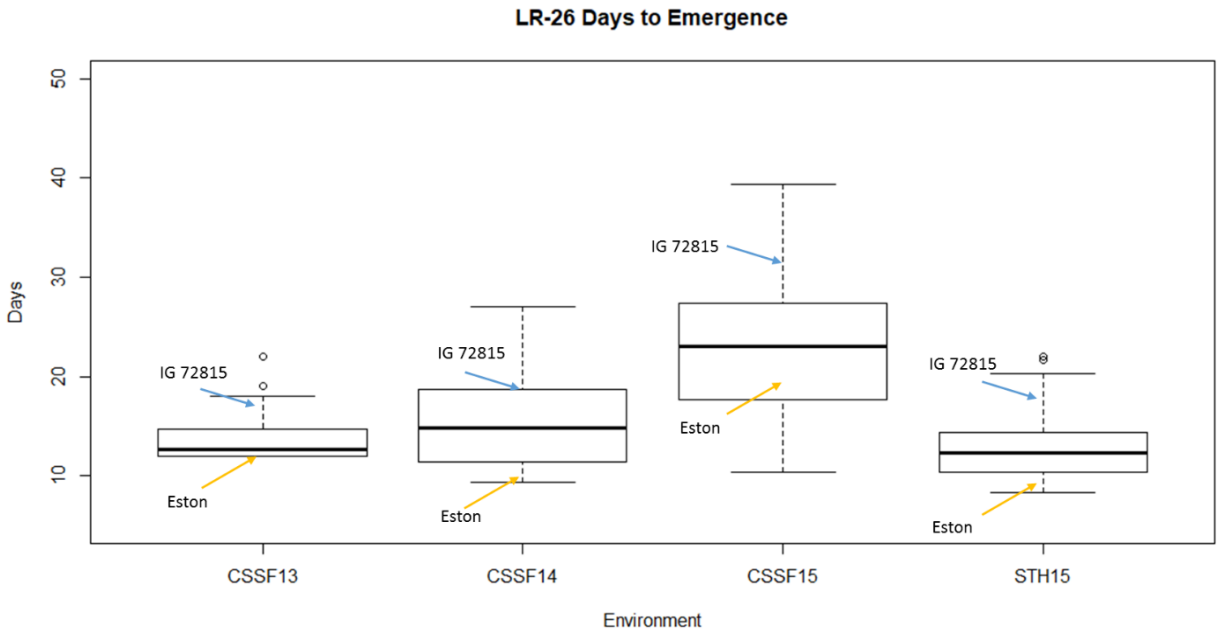


Figure 3.4 Distribution of days to emergence (DTE) in the LR-26 RILs (Eston x IG 72815) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of IG 72815 (Ler) are indicated with a blue arrow.

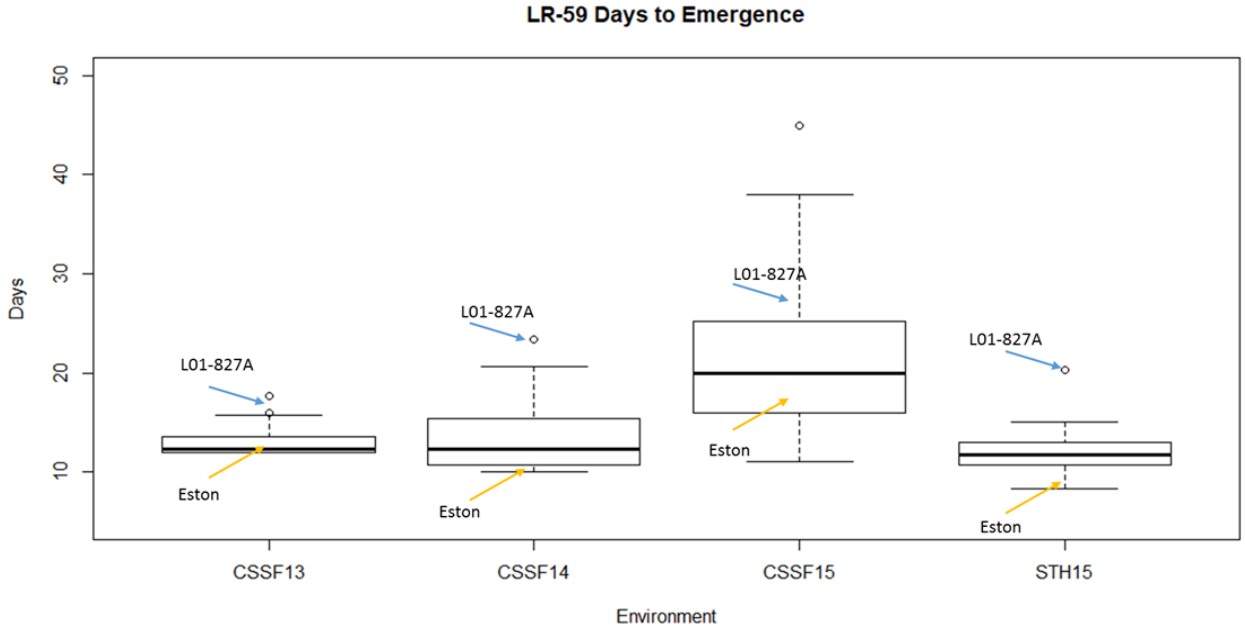


Figure 3.5 Distribution of days to emergence (DTE) in the LR-59 RILs (Eston x L01-827A) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of L01-827A (Ler) are indicated with a blue arrow.

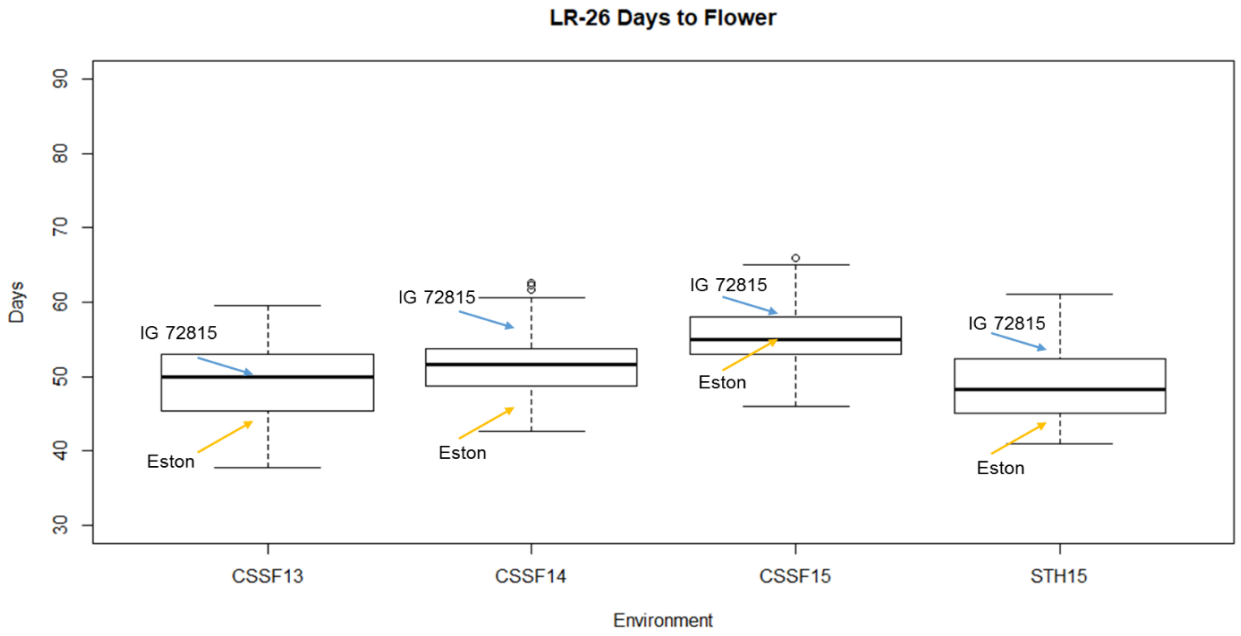


Figure 3.6 Distributions of days to flower (DTF) in the LR-26 RILs (Eston x IG 72815) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of IG 72815 (Ler) are indicated with a blue arrow.

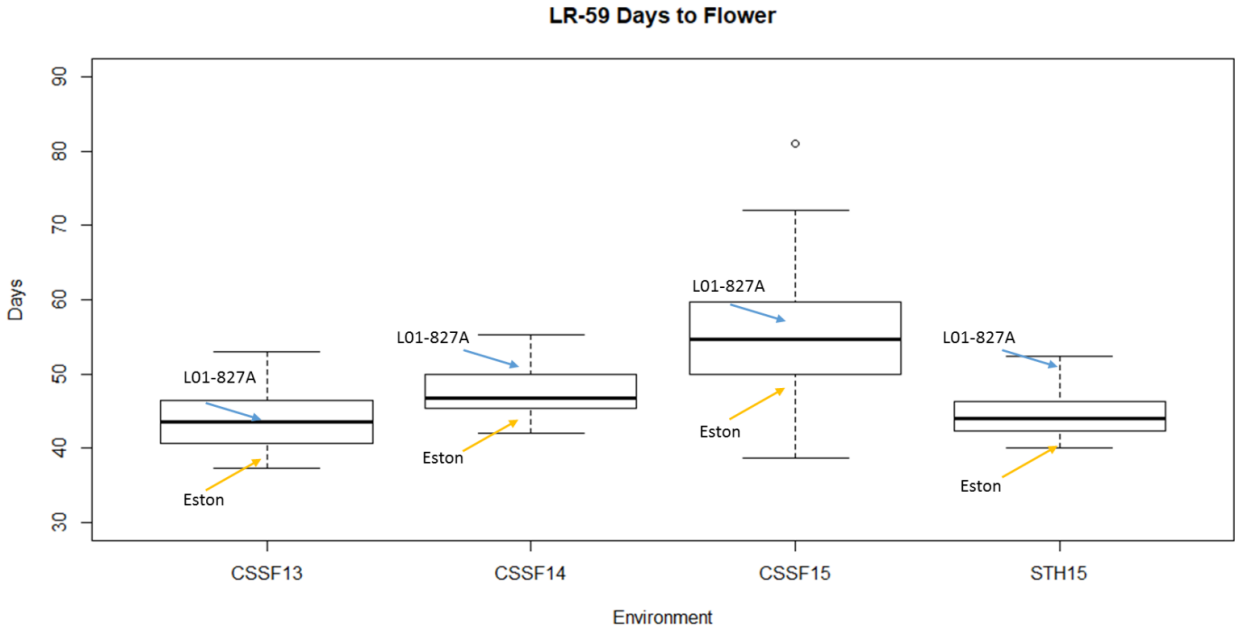


Figure 3.7 Distributions of days to flower (DTF) in the LR-59 RILs (Eston x L01-827A) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of L01-827A (Ler) are indicated with a blue arrow.

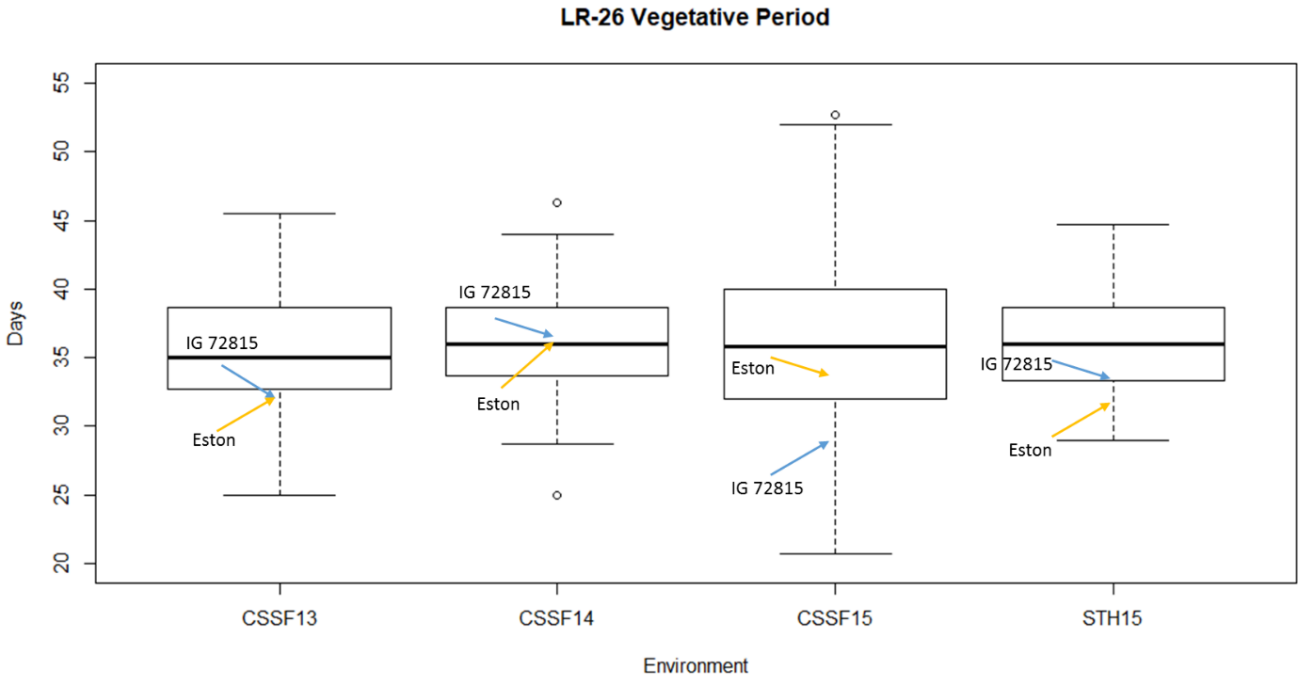


Figure 3.8 Distributions of vegetative period (VP) in the LR-26 RILs (Eston x IG 72815) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of IG 72815 (Ler) are indicated with a blue arrow.

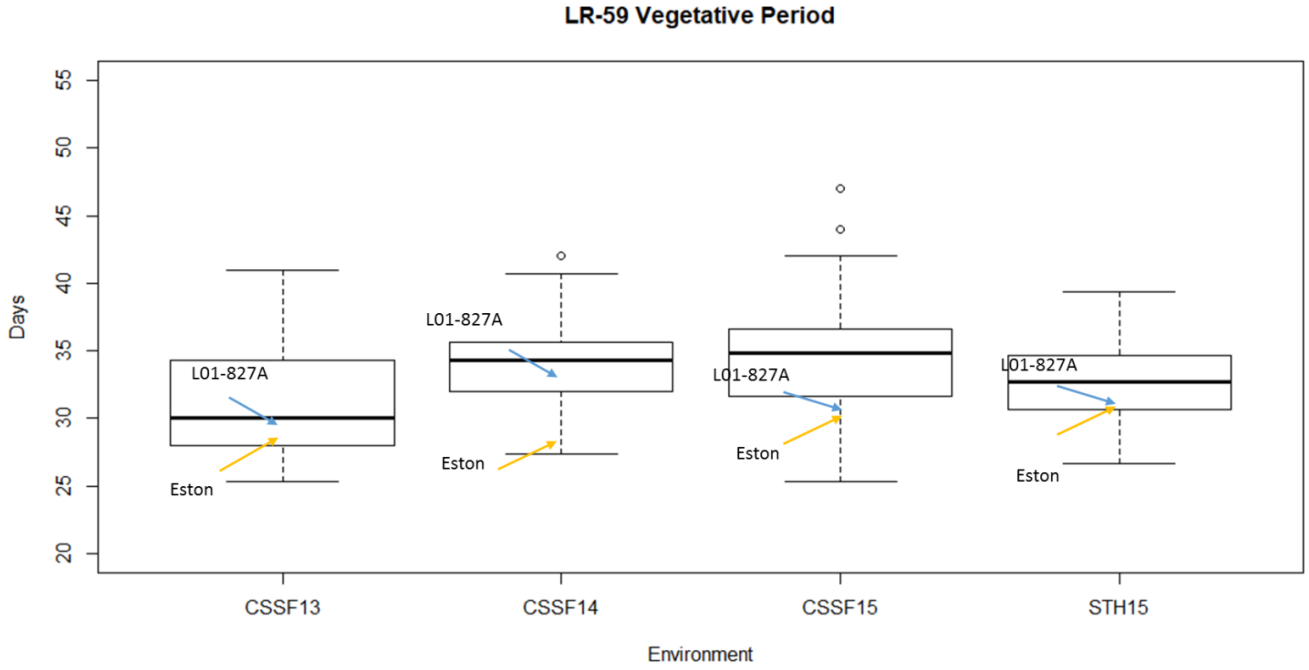


Figure 3.9 Distributions of vegetative period (VP) in the LR-59 RILs (Eston x L01-827A) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of L01-827A (Ler) are indicated with a blue arrow.

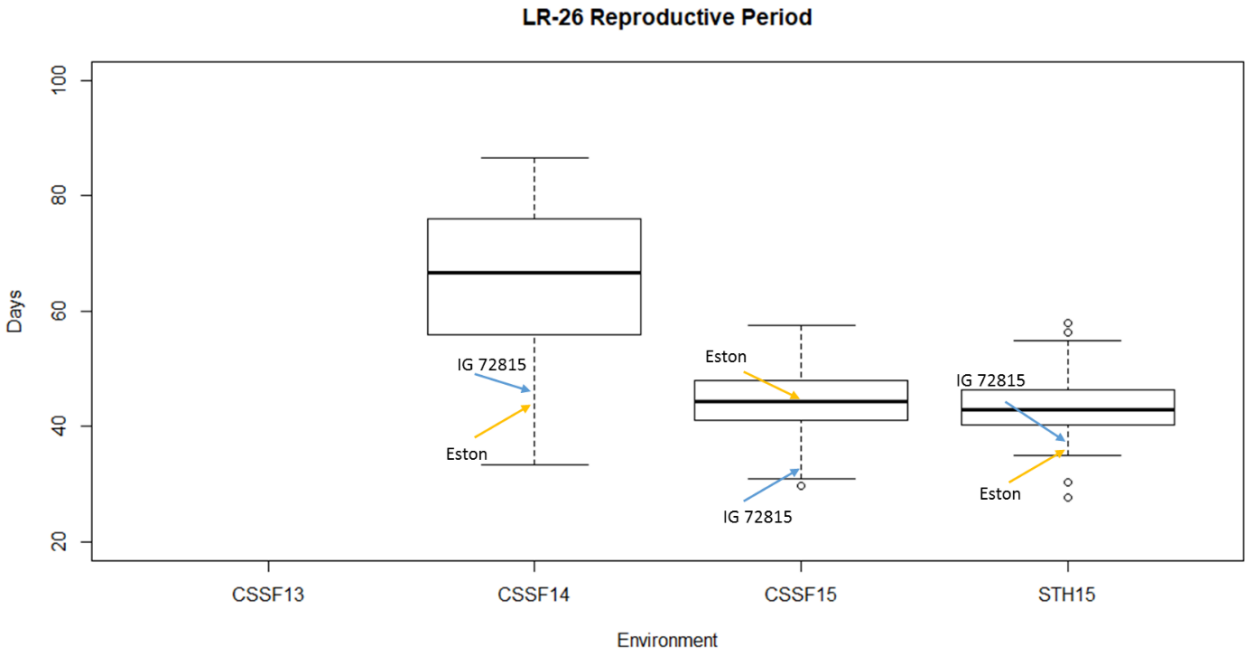


Figure 3.10 Distributions of reproductive period (RP) in the LR-26 RILs (Eston x IG 72815) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of IG 72815 (Ler) are indicated with a blue arrow.

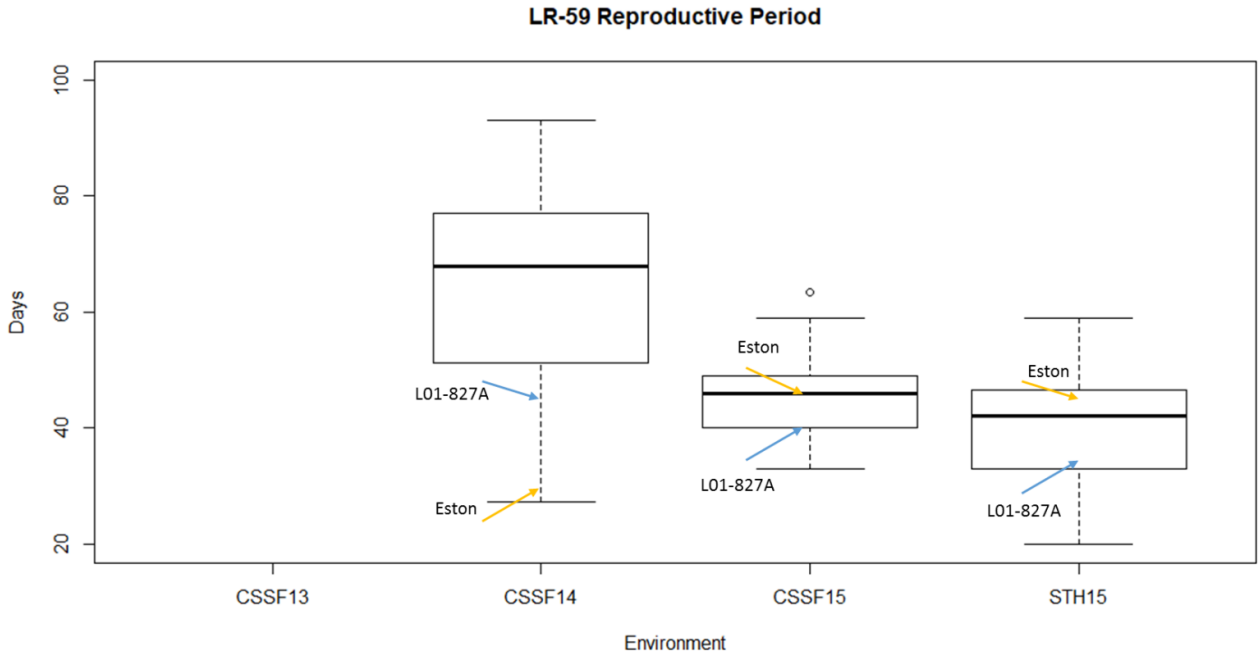


Figure 3.11 Distributions of reproductive period (RP) in the LR-59 RILs (Eston x L01-827A) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of L01-827A (Ler) are indicated with a blue arrow.

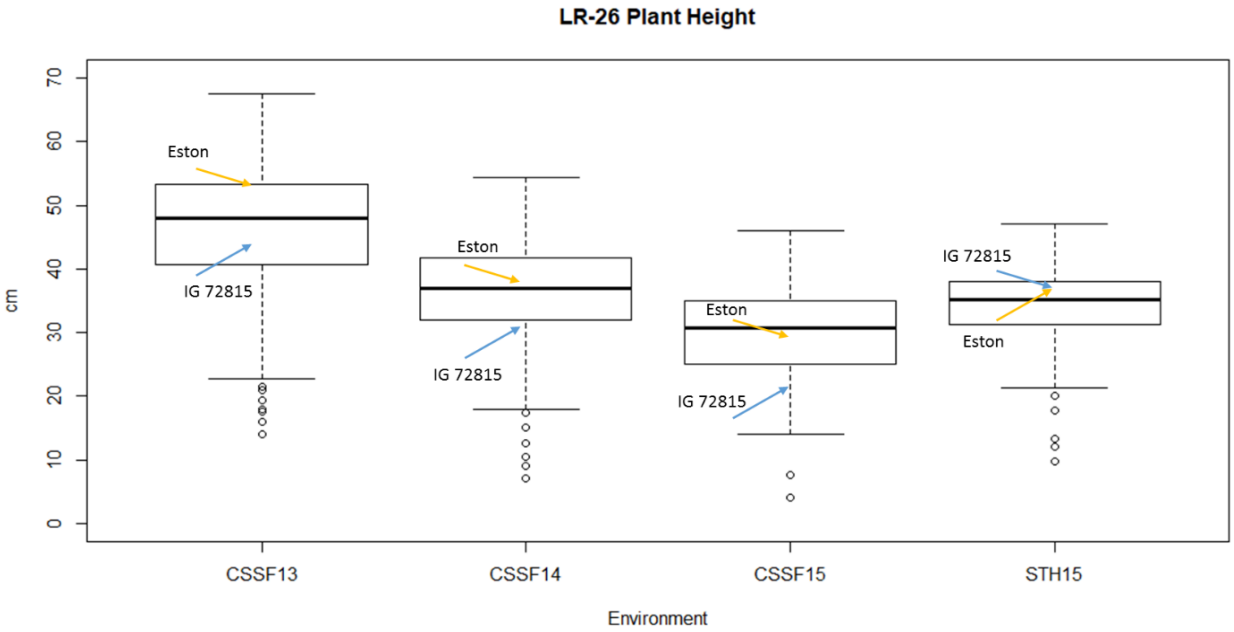


Figure 3.12 Distributions of plant height (PH) in the LR-26 RILs (Eston x IG 72815) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of IG 72815 (Ler) are indicated with a blue arrow.

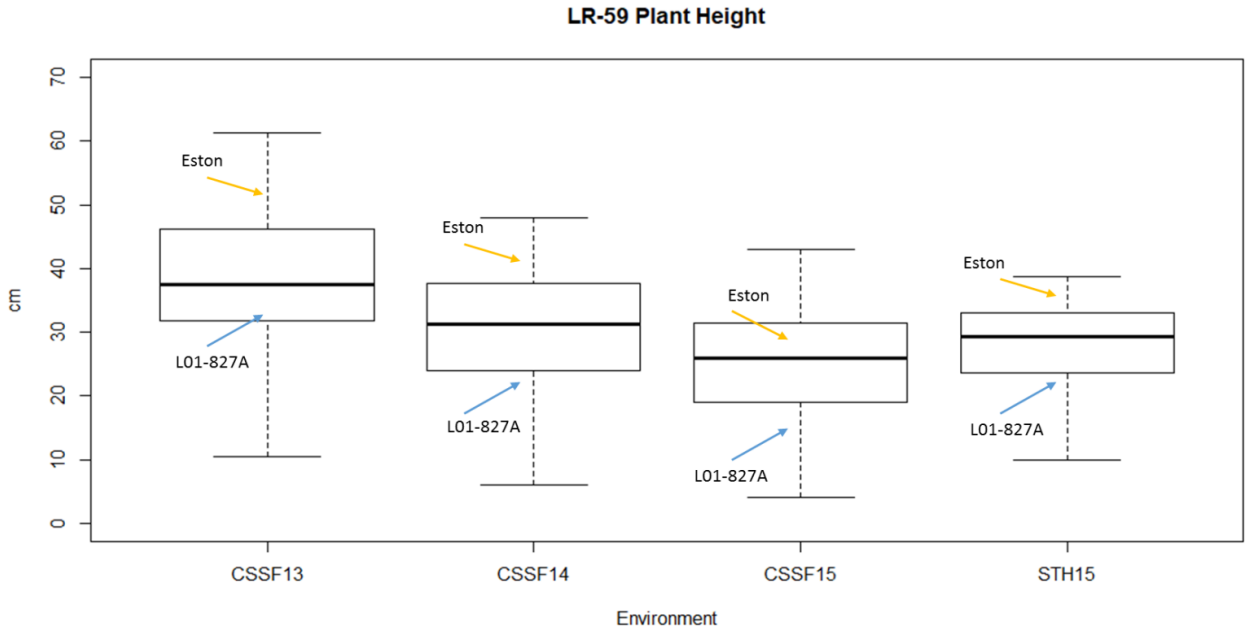


Figure 3.13 Distributions of plant height (PH) in the LR-59 RILs (Eston x L01-827A) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of L01-827A (Ler) are indicated with a blue arrow.

3.3.3 Estimation of variance components and broad sense heritability of agronomic quantitative traits

The estimated variance components and heritability of agronomic traits of the RILs from each population is summarized in Table 3.6 and 3.7, respectively. Among all the agronomic traits, RP had the lowest heritability (0.42 and 0.17 in LR-26 and LR59, respectively) (Table 3.6 and 3.7). The highest broad sense heritabilities were found for PH (0.80 and 0.94 in LR-26 and LR59, respectively) (Table 3.6 and 3.7) and DTF (0.79 and 0.83 in LR-26 and LR59, respectively) (Table 3.6 and 3.7).

3.3.4 Inheritance of flower colour and pod dehiscence

Flower colour in lentil has been described as extremely variable (Ladizinsky, 1979) with many different combinations. However, there were only two types of flower colour observed in this study, and the segregation of flower colour fitted a single locus model in both LR-26 and LR-59 given the results of the Chi^2 test (Tables 3.8; $P>0.05$).

Pod dehiscence segregated in a 1:1 ratio in LR-26 (Table 3.7; p-value=0.44). However, in LR-59, the segregation was distorted with more than 80% of the genotypes being resistant (Table 3.9). A two-gene model for pod dehiscence was further tested and showed that in LR-59, the segregation of pod dehiscence could fit a 1:3 segregation (p=0.12); while in LR-26, this two-loci model hypothesis was rejected (Table 3.8).

Table 3.6 Estimates of variance components and broad sense heritability of all agronomic quantitative traits in the LR-26 RILs (Eston x IG 72815).

Variance component	Days to emergence[J]	Days to flower	Vegetative period	Reproductive period	Plant height
σ^2e	1.88	14.56	0.12	165.78	49.81
σ^2g	4.82	9.40	4.92	13.888	30.00
σ^2g*e	1.40	3.82	6.16	34.73	15.67
σ^2p	6.30	11.89	7.97	32.98	37.69
H^2	0.77	0.79	0.62	0.42	0.80

σ^2e: environmental contribution to total phenotypic variation; σ^2g: total genetic variance; σ^2g*e: variance of genetic by environment interaction; σ^2p: total phenotypic variation; H^2: broad sense heritability; J: Days to emergence from CSFL15 of both LR-26 and LR59 was removed from the estimation of variance components and broad sense heritability

Table 3.7 Estimates of variance components and broad sense heritability of all agronomic quantitative traits in the LR-59 RILs (Eston x L01-827A).

Variance component	Days to emergence[J]	Days to flower	Vegetative period	Reproductive period	Plant height
σ^2e	0.38	3.21	2.48	5.50	22.86
σ^2g	0.44	5.18	4.48	3.69	53.35
σ^2g*e	1.36	1.52	3.93	28.74	0.25
σ^2p	0.11	6.22	6.34	21.51	56.96
H^2	0.28	0.83	0.71	0.17	0.94

σ^2e: environmental contribution to total phenotypic variation; σ^2g: total genetic variance; σ^2g*e: variance of genetic by environment interaction; σ^2p: total phenotypic variation; H^2: broad sense heritability; J: Days to emergence from CSFL15 of both LR-26 and LR59 was removed from the estimation of variance components and broad sense heritability

Table 3.8 Chi-square test of segregation ratios of mono-locus control for morphologic traits among RILs of LR-26 and LR-59.

	Flower colour of LR-26		Pod dehiscence of LR-26		Flower colour of LR-59		Pod dehiscence of LR-59	
	Non- purple	Purple	Resistant	Shattering	Non- purple	Purple	Resistant	Shattering
Source	Lcu	Ler	Lcu	Ler	Lcu	Ler	Lcu	Ler
Expected	86	86	82.5	82.5	32.5	32.5	33	33
Observed	92	80	76	89	28	37	55	11
X^2	0.84		0.59		1.25		29.33	
Probability	0.36		0.44		0.26		<0.0001	

X^2: Chi square test value. Both traits were tested at 1:1 ratio, all tests had a degree of freedom of 1

Table 3.9 Chi-square test of segregation ratios for a two loci control model on pod dehiscence among RILs of LR-26 and LR-59.

	Pod dehiscence of LR-26		Pod dehiscence in LR-59	
	Resistant	Shattering	Resistant	Shattering
Source	Lcu	Ler	Lcu	Ler
Expected	123.75	41.25	49.5	16.5
Observed	76	89	55	11
X^2	74.93		2.44	
Probability	<0.0001		0.12	

X^2: Chi square test value. Segregation on pod dehiscence was tested at 3:1 ratio, tests were under a degree of freedom of 1

3.4 Discussion

The *in situ* phenotypic evaluation of pre-breeding material is crucial for the utilization of exotic genetic resources (Sharma et al., 2013). Through conducting field trials and identifying phenotypic variation in LR-26 and LR-59, this study helps elucidate the potential challenges and opportunities of CWI in lentil. Overall, a broadening of the genetic base from hybridization between primary and tertiary genepools occurred in both LR-26 and LR-59 resulting in a high level of variation for several traits of agronomic importance.

In both LR-26 and LR-59, the emergence time of most of the RILs was later than the Lcu parent (Figures 3.4 and 3.5). Seed dormancy has been reported to be related to the hard seed coat of Ler (Ladizinsky, 1985b), but all the seeds in the trials were scarified before sowing to remove dormancy due to water-impermeability. The results suggest that there are sources other than hard seed coat that are delaying germination in the RILs. The DTE of both populations skewed towards earlier emergence, however, which might be due to a higher chance of the earlier emerging lines being retained during population development. Generally, the smaller population, LR-59, showed higher skewness in DTE than did LR-26, and almost all LR-59 RILs had an earlier emergence date than the Ler parent. This was probably because the later population, LR-26, was more carefully developed to maintain a large population size and to avoid selection during cycles of selfing (Tullu et al. 2013). The impact of cutworm invasion on emergence time biased the DTE of both LR-26 and LR-59 in CSSF15 (Table 3.5), therefore, this specific environment was removed from the analysis of DTE.

The time from seeding to first flower has been highly correlated to yield potential in lentil (Tambal et al., 2000) and is a very important indicator of adaptation. The variation in DTF would be affected by both seed emergence and plant flowering genetics, while VP, which is the difference between DTE and DTF, represents the

required period for individual genotype to flower post emergence. Among all the phenological traits in the RILs examined, low levels of variability were observed in DTF across the tested environments. DTF was also less variable within a given environment compared to other phenological traits in both populations. At the same time, VP was more stable compared to DTF within both populations. That is, the flowering time of Eston was consistently in late June to early July, while both IG 72815 and L01-827A would start to flower later, around mid to late July. Therefore, although the required time from seed emergence to flowering was similar between the parents, the DTF of Lcu was earlier than Ler in the tested environments because of later emergence time of the wild parents. Lentil is a quantitative long-day-species and flowering is influenced by both temperature and day-length during the vegetative period (Erskine et al., 1990; Roberts et al., 1986; Summerfield et al., 1985). The physiological control of lentil flowering relies on a certain thermal-photoperiod condition based on the specific time of year. This result indicates that the effect of the environment plays a significant role in VP among the interspecific RILs. Thus, the variation of VP in the RILs may better indicate the required condition for the individuals to flower in the tested environments of this study than does DTF.

Among all the agronomic traits recorded, RP was the least stable with a high environment component for the variation (Table 3.6). Since plant maturation is generally highly affected by temperature and humidity during pod development (Delouche, 1980), the high impact of environment on RP not only existed among the interspecific RILs but also can be found among the parents across environments; especially across years. Also, the indeterminate nature of lentils as well as the required environmental criteria of humidity and temperature for each genotype made the rating of exact maturation point very challenging under field conditions. The standard deviation (SD) of RP at CSSF14 was relatively large among tested environments in both LR-26 (SD = 16.43) and LR-59 (SD = 10.49). This overall delayed maturation was possibly

due to a cooler and more humid weather condition (Appendix A) during the reproductive stage in 2014 that resulted in an overall delayed RP compared to 2015 for both LR-26 and LR-59 (Figures 3.10 and 3.11).

The plant height has been determined to be a polygenic trait among various crop species and has been described to be one of the more heritable of the polygenic traits as reviewed in Fernandez et al. (2008). Although lentils are typically short, PH still has been reported to be highly variable from different lentil crosses (Hadda et al., 1982; Tullu et al., 2008). And Singh et al. (2017) reported higher variability in PH compared to other agronomic traits in Lcu x Ler derived inbred populations. In lentil, PH can be an indicator of plant structure and is very critical to the manageability and biomass of the plant (Tullu et al., 2001). However, in this study, the growth habit (erectness of the stem) was not evaluated. While the growth habit also segregated in both populations, the rating of erectness in this study would have been very difficult as the interspecific RILs were bagged in white mesh bags after flowering to avoid seed dispersal from pod shattering nature of the wild. Like other agronomic traits, PH also showed a large range of phenotypes in both LR-26 and LR-59 (Figures 3.12 and 3.13). In a related study of LR-26 from a greenhouse trial (Tullu et al., 2013), PH also showed high levels of variability and transgressive segregation. Although PH doesn't directly represent the growth habit and manageability in this study, it may still indicate the vigor of the RILs.

In this study, both LR-26 and LR-59 were derived from Lcu x Ler interspecific genomic backgrounds, and the phenotypic values were collected from the same environmental trials. Therefore, the phenotypic data of each trait from both populations were combined in this section for the estimation of heritability. Since the broad sense heritability was estimated from RILs of which the dominance and epistatic effects should be fixed, therefore the results can be compared with narrow sense heritability reported from related studies.

Overall, DTF and PH had the highest broad sense heritability in these two populations, while RP had the lowest (Table 3.6 and Table 3.7). Although RP had a low heritability of 0.52 (Table 3.6), it is noteworthy that there was no significant genotypic effect on RP at CSSF14 in LR-26 and only a slightly significant genotypic effect at the same environment in LR-59 (Tables 3.4 and 3.5). Both PH and DTF had a similarly high heritability (0.87 and 0.88, respectively) in these RILs. In a South Asian field-based study (Hamdi et al., 1991), the broad sense heritability of PH (0.90) and DTF (0.97) in lentil showed a similar higher level as the RILs in this study. In other lentil studies, PH displayed a higher narrow-sense heritability (0.65) than DTF (0.31) (Tullu et al., 2008).

In a study of a Lcu x Ler and a Lcu x *L. orientalis* derived populations by Singh et al. (2017), the heritability of days to flowering, days to maturity and plant height, were all reported to have narrow sense heritability estimates above 0.8 (h^2 = 0.86, 0.85, 0.83, respectively) based on data from both F_3 and F_4 generations. However, their study was conducted in one year of each generation with only two sites (Singh et al., 2017), therefore, likely experienced less of an environmental component compared to this study.

Other than the above quantitative variables, there were two segregating morphological traits, flower colour and pod dehiscence, that were characterized in the two populations. In lentil, the colour of flower has a wide range of variability among cultivated species and the wild species (Ladizinsky, 1979). The results in this study agreed with the conclusion of Tullu et al. (2013) that a single gene controls purple versus non-purple wing colour of the flowers in LR-26. A related study based on several F_2 populations derived from several *Lens spp.* interspecific crosses across three genepools (including Lcu x Ler) also concluded a single dominant gene underlies the purple flower colour of lentil (Singh et al., 2014b).

Another important morphological trait segregating in the interspecific RILs was pod dehiscence. In lentil, the selection of shattering-resistant mutants is a signature of

domestication (Hoffman et al., 1988) and an important factor in pre-breeding. Pod dehiscence is the natural seed dispersal mechanism at maturation, and a lack of dehiscence is typically included as one of the domestication syndrome traits among field crop species. There are two main physiological forces controlling this mechanism. In the first, medial pod valves separate due to reduced pod wall-binding strength from the replum, resulting in pod shattering (Dong and Wang, 2015); this force and pod abscission usually occur at the same time. The second force is a lateral outward force that results in the formation of a spiral coil of the pod walls; this is the one characterized in the populations studied here. These two forces generate the separation layer and lignified layer of the dehiscent zone. According to Ladizinsky (1979), pod dehiscence is controlled by a single dominant gene. Tullu et al. (2013) and Singh et al. (2014b) both reported this trait fit a single-locus control model based on Lcu x Ler derived populations. In LR-26, the 1:1 segregation result fits with the single-gene model expected based on these previous reports (Table 3.8). However, in LR-59, the population was distorted with more than 80% of the genotypes being non-dehiscent, and the result better fitted a two-loci model (Tables 3.9). Considering both LR-59 to LR-26 were derived from Lcu x Ler crosses, however, the difference in results is more likely due to management during population development than an actual second gene. The population size of LR-59 declined from more than 180 lines in the F_2 generation to only 66 RILs in the current study. Other than the reported fertility issue (Fiala et al., 2009), conscious/ unconscious selection against pod shattering type plants could also have contributed to the distortion around this gene.

Looking at the frequency distributions presented here, phenotypic distortion can be found for many of the agronomic traits. Such phenomena may be attributed to two major forces. First, the RILs may have undergone environmental and possibly unconscious selection during population development. The impact of these kinds of selection are also reflected in the shrinking size of both populations as they were selfed

from a large F_2 population to F_7-derived RILs, especially in LR-59. The second possible cause of segregation distortion is the divergent genetic background and possible difference in genome structure between the domesticated parent and the tertiary genepool species parent. Since the populations were derived from such a diverse cross, possible incompatibility and genome rearrangement between the parental species may have led to an aberrant pairing during meiosis resulting in unequal segregation in the interspecific progeny.

Phenotypic heterogeneity was observed within several RILs in both interspecific populations, resulting in non-uniformity within RILs of some phenotypes. Since the population was derived from the F_7 generation using single seed descent, the expected heterozygosity should be approximately 1.5%. This phenotypic heterogeneity may be attributed to the aberrant chromosome paring during meiosis because of the evolutionary divergence between two parental species (Ladizinsky et al., 1985; Gujaria-Verma et al., 2014). Segregation distortion and possible genome rearrangements have been reported in other Lcu x Ler populations based on a limited number of isozyme and morphological markers (Zamir and Tadmor, 1986; Tadmor et al., 1987). A dense genotyping of one of these populations would give a clearer view of the consequences of genome introgression between these two species (see Chapter 5). Generally, a significant genotypic effect was found in LR-26 for all the recorded traits among the tested environments, except for DTE in CSSF15 due to the cutworm invasion, and for RP in CSSF14 which also had a high environmental impact. Therefore, it should be possible to combine the results from this chapter with the LR-26 genotyping results in Chapter 5 to detect the Ler introgression regions associated with these phenotypes (see Chapter 6).

Prologue to Chapter 4

Genetic erosion from population bottlenecks has brought up the challenge of genetic vulnerability among major crop species. Natural genetic resources such as landraces, crop progenitor species and distant wild relative species are potential sources of some desirable variations and traits. They can bring along unexpected or negative variability, too.

In both LR-26 and LR-59, all the traits tested showed significant genotypic effect within a given single environment, and a significant G x E interaction was found across growing environments. Additionally, transgressive segregation was observed for several traits of agronomic importance. It should be noted that along with high variability, however, comes undesirable phenotypes resulting in poor performance. These phenotypes include the re-introduction of wild traits such as pod dehiscence, shorter plant type and later emergence than the cultivated parent. Plant breeders are not only interested in agronomic traits, however, and to further evaluate the effects of CWI, other important traits such as seed quality were investigated and are reported in Chapter 4.

CHAPTER 4. ASSESSING THE IMPACT OF INTROGRESSION ON SELECTED SEED QUALITY TRAITS OF LENTIL IN THE INTERSPECIFIC RECOMBINANT INBRED LINE POPULATIONS LR-26 AND LR-59

4.1 Introduction

Seed visual quality characteristics, such as seed size, seed coat colour and cotyledon colour, are crucial for lentil market value, grading and exportability. During field trials (Chapter 3), three important seed visual characteristics, including seed size, hilum colour and cotyledon colour, were found to segregate among the LR-26 and LR-59 interspecific RILs.

In addition to visual quality characteristics, nutritional composition is an important seed quality trait in lentil. Among pulse crop species, sucrose and the raffinose family oligosaccharides (RFO) are the most abundant seed storage carbohydrates (Obendorf and Gorecki, 2012). While RFO in lentil seeds are considered as beneficial dietary fibre (Agil et al., 2013; Dwivedi et al., 2014; Johnson et al., 2013), a higher level of RFOs can also cause bloating because of indigestibility due to the lack of α-galactosidase in the monogastric gut, which may lead to a reluctance in consumption.

To boost pulse consumption, studies have been carried out aiming to reduce the total RFOs (TRFO) level of seed in some other grain legume species (Dierking and Bilyeu, 2008; Hitz et al., 2002; Jones et al., 1999; Yang et al., 2014). To assess this breeding approach in lentil, the first step was to explore the compositional variations among lentil genetic resources. Tahir et al. (2012) screened seed sucrose and TRFO concentration from an International Center for Agricultural Research in Dry Areas (ICARDA) *Lens* germplasm collection. They reported an overall lower level of these two storage carbohydrates among the *L. ervoides* (Ler) accessions in their collection. To follow up, a preliminary experiment was done to measure the concentrations of TRFO in the parents of LR-26 and LR-59, and in both parental combinations there were

contrasting levels (Appendix B). Based on the above evidence, this study was designed to investigate the segregation of these seed soluble carbohydrates in LR-26 and LR-59 along with the visual quality characteristics.

4.2 Materials and methods

4.2.1 Plant material

Two RIL populations, LR-26 and LR-59, were used in this study as described in section 3.2.1 of Chapter 3. All seed samples were obtained from the field trials as noted in Table 3.1 of Chapter 3. Seed samples were collected from bulk harvest of corresponding field trials. All plants of one plot were covered in one to several white mesh bags at flowering stage, depending on the number and size of plants in the plot. After maturation, all plants from one plot were collected inside a labeled paper bag. The labels indicated the genotype, entry number (referred to the replication) and environment of trial. Labeled paper bags were collected into canvas bags to be heat dried. The dried plants were then threshed, and seeds of each plot were cleaned and collected in a labeled paper envelope for further use.

4.2.2 Seed quality characterization

4.2.2.1 Quantitative traits

4.2.2.1.1 Estimation of thousand seed weight (TSW): Seed weight was used to represent the quantitative variation of seed size and recorded as weight, in grams, of a thousand seeds. Seed samples were cleaned and two hundred seeds from each plot were counted using an electronic seed counter (ESC-1, Agriculex Inc., Canada). Seeds were weighed on a balance and the mass was used to calculate TSW. This trait was calculated with seeds harvested from CSSF13, CSSF14, CSSF15 and STH15.

4.2.2.1.2. Estimation of seed sucrose concentration and seed total raffinose family oligosaccharides (TRFO) concentration: Samples were only tested from the genotypes which yielded more than 5 g of seeds from at least two replications at a location.

Approximately 2 g of bulked seeds of each genotype from each replication were weighed, and seed samples were milled to a fine flour using a cyclone sample mill (Udy Corporation, USA). The flour from seeds of the same plot was mixed homogeneously in a sealed bag prior to analysis. Each plot was treated as one biological replication. The analysis of sucrose and TRFOs concentration was done with the enzymatic hydrolysis method using a commercialized colorimetric assay kit (Raffinose/ Sucrose/ D-Glucose assay kit, Megazyme International, USA); where the concentration of different members of TRFO was determined as a group. Using the Megazyme kit, in principle, TRFOs in the seed flour would to be hydrolysed by α-galactosidase into D-galactose and sucrose and then sucrose would to be digested by invertase into D-glucose and D-fructose. The resulting D-glucose would react with glucose oxidase/peroxidase (GOPOD) reagent and be oxidized into D-gluconate and hydrogen peroxide. Hydrogen peroxide would be catalyzed by peroxidase into quinonimine (with red colour) by p-hydroxybenzoic acid (provided in the GOPOD buffer) and 4-aminoantipyrine (in the GOPOD reagent). The colour was then quantified at A510 nm using a spectrophotometer. In depth steps were as follows:

Before starting the assay, several buffers and working solutions were prepared. First the 50 mM sodium acetate buffer was prepared by adding 2.9 ml of glacial acetic acid to 900 ml of nano-pure water and adjusting to pH 4.5 using 1 M sodium hydroxide. The final volume of the 50 mM sodium acetate buffer was adjusted to 1 L and the prepared buffer was stored at 4°C. Then the working solution of invertase, and invertase and α-galactosidase enzyme mixture (stock solutions of enzymes provided in the kit) was prepared. The working solution of invertase was prepared by adding 1 ml of invertase stock solution to 12 ml of the sodium acetate buffer and mixed thoroughly. The working solution of enzyme mixture was prepared by adding 2 ml α-galactosidase stock solution and 2 ml of the invertase stock solution to 21 ml of the sodium acetate buffer and mixed thoroughly. Last, the GOPOD reagent and GOPOD reagent buffer

were prepared. The GOPOD reagent buffer was prepared by diluting 50 ml of the GOPOD reagent buffer stock solution (stock solution provided in the kit) to 1 L using nano-pure water. The GOPOD reagent was prepared by first adding 20 ml of GOPOD reagent buffer to the sealed bottle of freeze-dried GOPOD powder provided in the kit to dissolve the powder and then the content was transferred to the bottle containing the remainder of GOPOD reagent buffer. This buffer was covered with aluminum foil to be protected from light.

The assay was processed according to the manufacturer's instruction as below. For sugar extraction, 500 mg of seed flour of each sample was weighed into a centrifuge tube (50 ml) and 5 ml of 95% ethanol was added. Tubes were then incubated at 85 °C for 5 minutes to inactivate endogenous enzymes. Sodium acetate buffer was added to each tube to 50 ml. The samples were then mixed thoroughly and extracted over 15 minutes, 5 ml of the sample slurry was transfer to a new centrifuge tube, and 2 ml of chloroform was added to the slurry solution. The solution was vortexed for 15 seconds and then centrifuged at 1,000 xg for 10 minutes. Solution of the upper aqueous phase was used as the solution A of the assay. For each sample, three centrifuge tubes were prepared for glucose, sucrose and TRFOs, respectively; and tubes were labeled as Tube A, Tube B and Tube C, accordingly. For each tube, the assay mixture was prepared as listed: Tube A: 200 µl of solution A + 200 µl of 50 mM sodium acetate buffer. Tube B: 200 µl of solution A + 200 µl of invertase working solution. Tube C**:** 200 µl of solution A + 200 µl of enzyme mixture working solution of α-galactosidase and invertase. A reagent blank was prepared by mixing 400 µl of sodium acetate buffer and three tubes of glucose standard solution, each glucose standard solution was made by adding 100 µl of D-glucose standard stock (1 mg/ml in 0.2% benzoic acid) to 300 µl of sodium acetate buffer. All tubes were incubated at 50 °C for 20 minutes. After the incubation, 3 ml of GOPOD reagent was added to each tube. All tubes were then further incubated at 50 °C for 20 min. After the second incubation, 250 µl of solution was transferred to a flat-

bottom, transparent 96-well microplate to read the absorbance of all solutions against the reagent blank at A510 nm using a FLUOstar microplate reader (BMG Labtech, USA). Absorbance was divided as follows: ΔA = GOPOD absorbance of tube A for D-glucose; ΔB = GOPOD absorbance of tube B for D-glucose plus sucrose; ΔC = GOPOD absorbance of tube C for D-glucose plus sucrose plus TRFO. Sucrose and TRFO concentration were calculated using the following equations in mM/100 g of seed flour: Sucrose = (ΔB-ΔA) x F x 250 x 200 x 1/1000; TRFO = (ΔC-ΔB) x F x 250 x 200 x 1/1000, where F was a coefficient factor to convert absorbance into micromoles of glucose.

F was calculated as: F = 0.556 (μmoles of glucose)/GOPOD absorbance of glucose standard. The conversion of 50 ml of extract to 500 mg of sample was made by multiplying by 250 and the conversion from 500 mg to 100 g of sample was completed by multiplying by 200. The value was subsequently converted from μmoles to mmoles. The concentration of sucrose was further converted to grams to estimate sucrose concentration in the seeds by dividing by 1000 and then converted to moles by multiplying by the molecular mass of sucrose of 342.3. Both sucrose and TRFO concentration were calculated with seeds harvested from CSSF14, CSSF15 and STH15.

4.2.2.2 Qualitative seed quality traits

During field trials observations, hilum colour and cotyledon colour of seeds were found to be segregating into two categories of each trait among the interspecific RILs. The rating was based on the characteristics of the contrasting parents (Figure 4.1). The traits were evaluated by seeds of each genotype harvested in CSSF14 trial and further confirmed using seed in trials of CSSF15 and STH15.

Hilum colour: The hilum is the oval scar on the seed coat. The pigmentation level of the hilum may affect the market value of lentil. Hilum colour in the populations was classified based on the ring around the hilum scar as dark (or self-coloured, as the Lcu parent) or light (as the Ler parent).

Seed cotyledon colour: Cotyledon colour is very important for market classification of lentil. This trait can be characterized as yellow (as the Lcu parent) or red (as the Ler parent) in both populations.

Figure 4.1 Typical seed morphology represented by the two parental species. Left (I): Lcu (variety Eston), standard small, green type of lentil. Right (II): Typical Ler seeds, photo shows accession IG 72815.

4.2.3 Data analysis

All the data analysis was done using SAS software version 9.4 (SAS Inc., USA). The processing of quantitative and qualitative data analyses was the same as described in section 3.2.4 of Chapter 3. A Phi coefficient was calculated in Microsoft Excel 2016 using a 2×2 contingency table to estimate the possible correlation between the two binary traits, seed hilum colour and pod dehiscence.

4.3 Results

4.3.1 Variability in quantitative seed quality traits

The ANOVA F-test results for TSW, sucrose and TRFO concentrations from the multi-environment trials are summarized in Tables 4.1 and 4.2 for LR-26 and LR-59, respectively. The results showed that all traits had highly significant ($p<0.001$) genotypic effects in both populations. There were highly significant ($p<0.001$) environmental effects on TSW for both populations (Tables 4.1 and 4.2), and the environmental impact on sucrose was slightly significant ($p<0.05$, Table 4.1) in LR-26

but not significant (p= 0.06, Table 4.2) in LR-59. For TRFO, the environmental effect was not significant (p= 0.18, Table 4.1) in LR-26 but highly significant (p<0.001, Table 4.2) in LR-59. However, like the results for quantitative agronomic traits in these populations (Tables 3.2 and 3.3), these three seed quality traits showed highly significant (p<0.001) GxE interactions in both populations. Because of this, all three traits were further analyzed by individual environments. The F-test results of the ANOVAs for TSW, sucrose and TRFO concentrations from each environment are summarized in Tables 4.3 and 4.4, for LR- 26 and LR-59, respectively. The results showed that all three seed quality traits had highly significant (p<0.001) genotypic effect within each environment in both LR-26 and LR-59 (Tables 4.3 and 4.4).

4.3.2 Frequency distributions and variations in seed quality traits

In both populations, the two parents showed contrasting levels for all the seed quality traits examined. The variations segregated among the individual RILs in a quantitative manner (Figures 4.2 – 4.7). Parents of LR-26 (Figure 4.2) and parents of LR-59 (Figure 4.3) both showed contrasting seed size in all environments, and TSW segregated in an additive manner in both populations. For both populations, the RILs were all slightly skewed toward smaller seed size (Figures 4.2 and 4.3), and transgressive segregation was not evident for this trait. For sucrose and TRFO concentration, the parents of LR-26 had contrasting levels at all environments (Figures 4.4 and 4.6). In CSSF15, very limited seed was collected from the wild parent of LR-59 due to poor emergence after the cutworm invasion. So, in this specific environment, only one biological replication of L01-827A had enough seed for analysis. Parents of LR-59 had contrasting concentrations of sucrose and TRFOs in CSSF14 and STH15 (Figures 4.5 and 4.7). For both populations, the RILs skewed toward higher concentrations of these two storage carbohydrates (Figures 4.4 - 4.7).

Table 4.1 F-test results from an ANOVA of three seed quality traits recorded for LR-26 RILs grown in four environments in Saskatchewan, Canada.

	TSW		Sucrose		TRFO	
Effect	df	F Value	df	F Value	df	F Value
Genotype	173	26.45***	158	7.38***	156	1.40**
Environment	3	18.84***	2	6.88*	2	1.18ns
Replication|Environment	8	9.47**	6	4.94***	6	0.42ns
Genotype*Environment	500	2.32***	244	3.05***	230	1.78***
C.V. (%)	35.03		24.80		37.06	

***: significant at P< 0.001; **: significant at P< 0.01; *: significant at P< 0.05; ns: not significant; df: degree of freedom; C.V.: coefficient of variation; TSW: Thousand seed weight; TRFO: Total raffinose family oligosaccharide

Table 4.2 F-test results from an ANOVA of three seed quality traits recorded for LR-59 RILs grown in four environments in Saskatchewan, Canada.

	TSW		Sucrose		TRFO	
Effect	df	F Value	df	F Value	df	F Value
Genotype	67	43.75***	65	2.85***	63	1.73*
Environment	3	24.47***	2	3.18ns	2	49.77***
Replication|Environment	8	3.12*	6	3.06**	6	1.12ns
Genotype*Environment	197	1.70***	87	4.12***	75	4.82***
C.V. (%)	30.98		33.40		24.65	

***: significant at P< 0.001; **: significant at P< 0.01; *: significant at P< 0.05; ns: not significant; df: degree of freedom; C.V.: coefficient of variation; TSW: Thousand seed weight; TRFO: Total raffinose family oligosaccharide

Table 4.3 F-test results of an ANOVA of the quantitative seed quality traits recorded for LR-26 RILs at each of four growing environments.

Environment	Source	TSW		Sucrose		TRFO	
		df	F Value	df	F Value	df	F Value
CSSF13	Genotype	172	52.50***	--	--	--	--
	Replication	2	4.86**	--	--	--	--
CSSF14	Genotype	170	44.76***	149	51.30***	133	14.57***
	Replication	2	3.11*	2	4.16*	2	0.28 ns
CSSF15	Genotype	164	5.59***	148	15.94***	146	10.89***
	Replication	2	1.22ns	2	10.97***	2	5.37*
STH15	Genotype	167	45.31***	105	11.26***	107	10.03***
	Replication	2	24.14***	2	1.74ns	2	0.52 ns

***: significant at P< 0.001; **: significant at P< 0.01; *: significant at P< 0.05; ns: not significant; df: degree of freedom; C.V.: coefficient of variation; CSSF13: Crop Science seed farm at 2013; CSSF14: Crop Science seed farm at 2014; CSSF15: Crop Science seed farm at 2015; STH15: Sutherland experiment farm at 2015; TSW: Thousand seed weight; TRFO: Total raffinose family oligosaccharide

Table 4.4 F-test results of an ANOVA of the quantitative seed quality traits recorded for LR-59 RILs at each of four growing environments.

Environment	Source	TSW		Sucrose		TRFO	
		df	F Value	df	F Value	df	F Value
CSSF13	Genotype	65	45.33***	--	--	--	--
	Replication	2	2.58ns	--	--	--	--
CSSF14	Genotype	67	26.28***	51	88.07***	42	23.10***
	Replication	2	0.32ns	2	0.27 ns	2	0.58ns
CSSF15	Genotype	65	8.24***	39	7.05***	39	3.04***
	Replication	2	2.01ns	2	4.36**	2	0.73 ns
STH15	Genotype	65	17.85***	63	7.08***	61	5.40***
	Replication	2	6.30**	2	2.24 ns	2	1.23 ns

***: significant at P< 0.001; **: significant at P< 0.01; *: significant at P< 0.05; ns: not significant; df: degree of freedom; C.V.: coefficient of variation; CSSF13: Crop Science seed farm at 2013; CSSF14: Crop Science seed farm at 2014; CSSF15: Crop Science seed farm at 2015; STH15: Sutherland experiment farm at 2015; TSW: Thousand seed weight; TRFO: Total raffinose family oligosaccharides

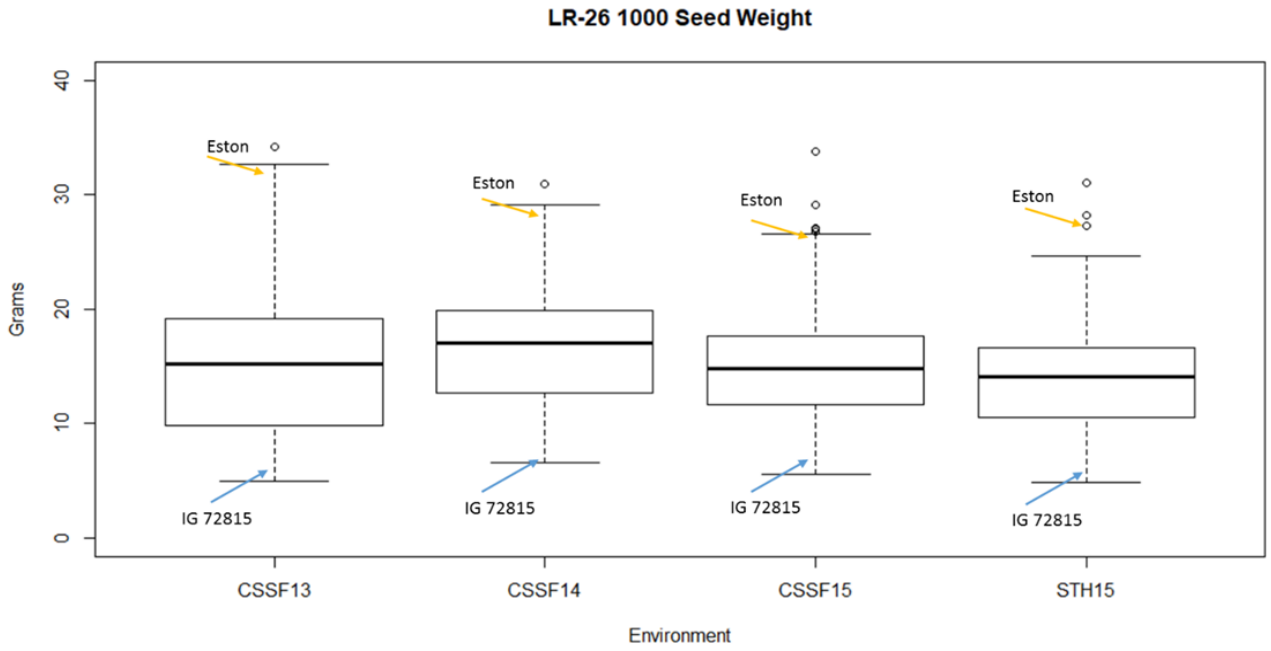


Figure 4.2 Distribution of thousand seed weight (TSW) in the LR-26 RILs (Eston x IG 72815) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of IG 72815 (Ler) are indicated with a blue arrow

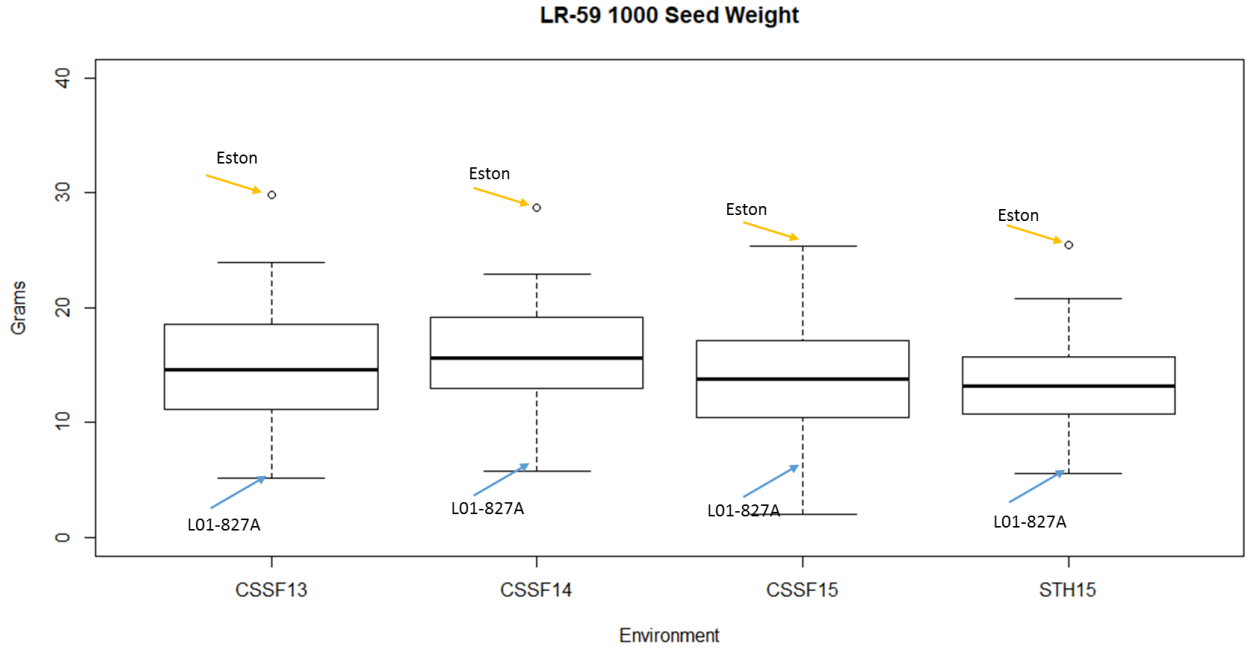


Figure 4.3 Distribution of thousand seed weight (TSW) in the LR-59 RILs (Eston x L01-827A) in four environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of L01-827A (Ler) are indicated with a blue arrow.

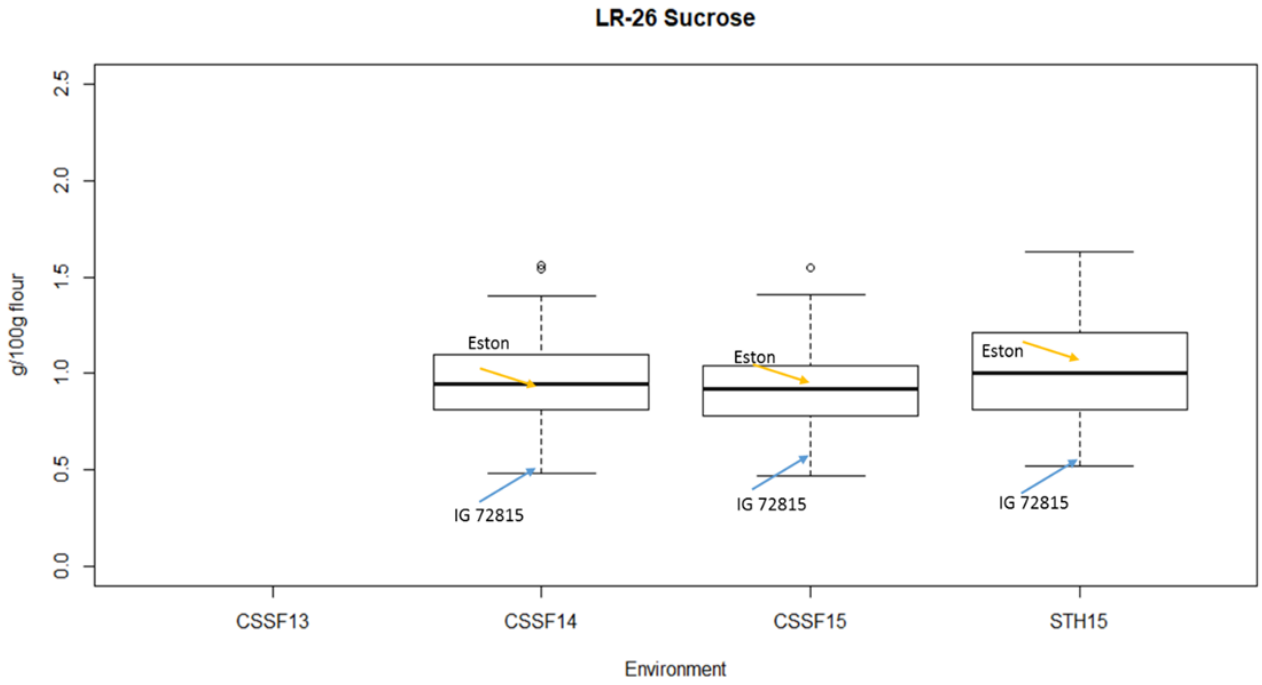


Figure 4.4 Distribution of sucrose concentration in the LR-26 RILs (Eston x IG 72815) in three environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of IG 72815 (Ler) are indicated with a blue arrow.

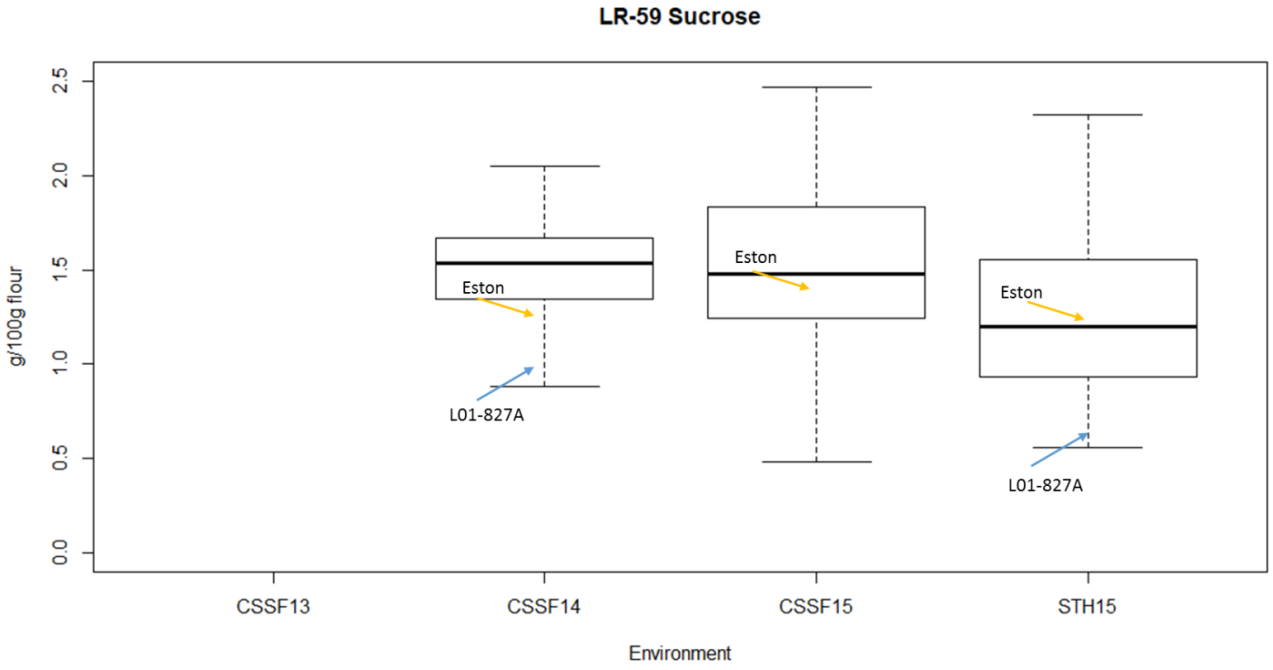


Figure 4.5 Distribution of sucrose concentration in the LR-59 RILs (Eston x L01-827A) in three environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of L01-827A (Ler) are indicated with a blue arrow.

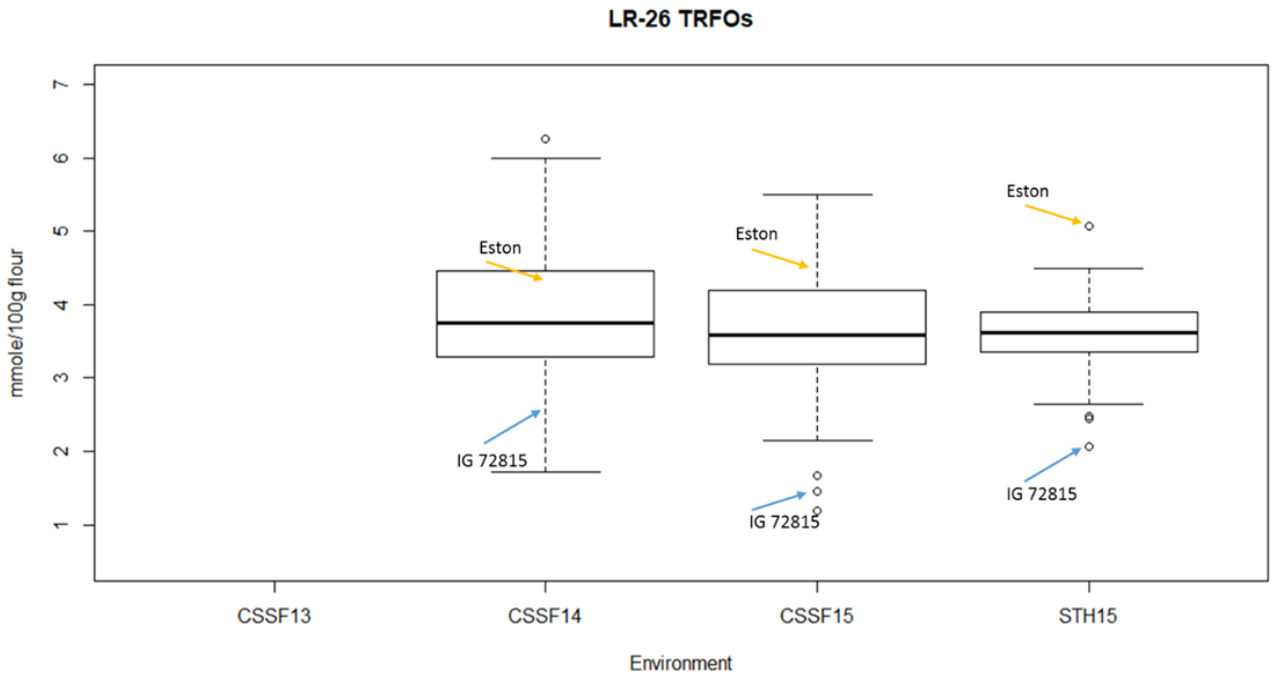


Figure 4.6 Distribution of total raffinose family oligosaccharides (TRFO) concentration in the LR-26 RILs (Eston x IG 72815) in three environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of IG 72815 (Ler) are indicated with a blue arrow.

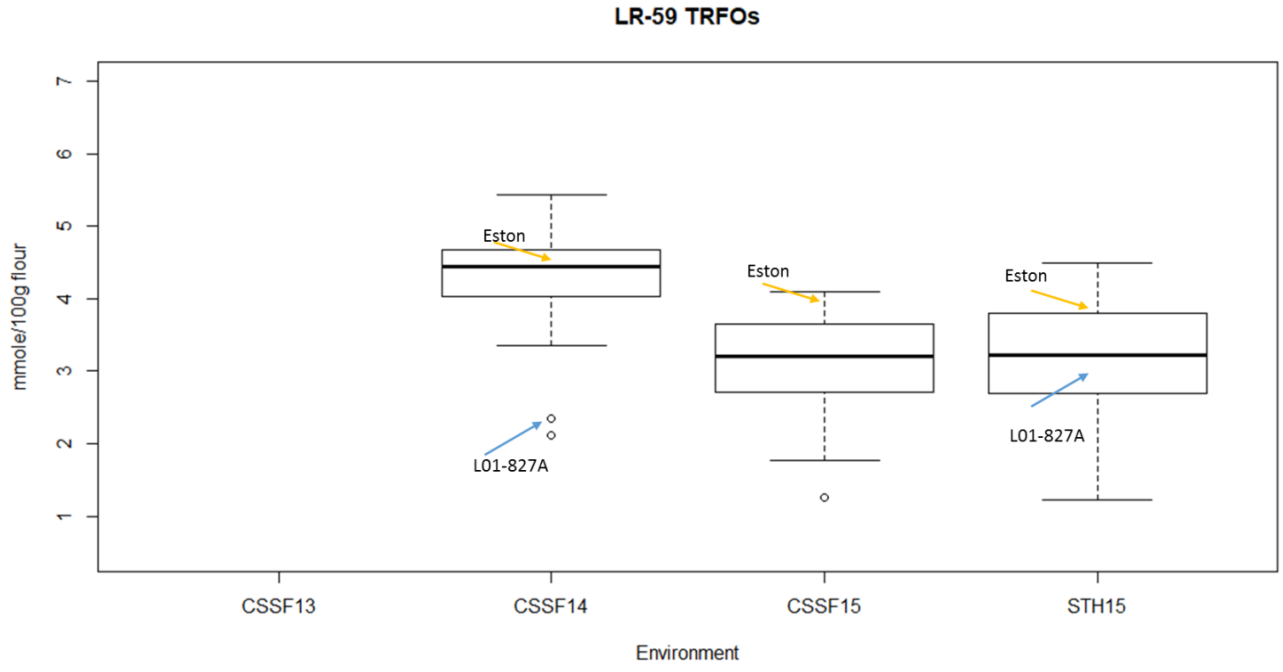


Figure 4.7 Distribution of total raffinose family oligosaccharides (TRFO) concentration in the LR-59 RILs (Eston x L01-827A) in three environments. The mean values of Eston (Lcu) are indicated with a yellow arrow; the mean values of L01-827A (Ler) are indicated with a blue arrow.

4.3.3 Estimation of variance components and broad sense heritability for seed quality traits

Table 4.5 and Table 4.6 show the estimated variance components and heritability estimates for the three seed quality traits investigated in LR-26 and LR-59, respectively. The broad sense heritability of TSW (0.96 and 0.98 in LR-26 and LR-59, respectively) (Table 4.5 and 4.6) was the highest among all the traits in this study including the agronomic ones (Table 3.6 and Table 3.7). Sucrose also had a high to medium broad sense heritability in both populations (0.89 and 0.69 in LR-26 and LR-59, respectively) (Table 4.5 and 4.6) while the heritability of TRFOs (0. 72 and 0.42 in LR-26 and LR-59, respectively) (Table 4.5 and 4.6) was relatively lower among the seed quality traits.

4.3.4 Inheritance of cotyledon and hilum colour

Two visual seed quality traits, cotyledon colour and hilum colour, were contrasting between the parental species in this study. Table 4.7 summarizes the segregation of these two traits among the RILs of LR-26 and LR-59. The two cotyledon colours segregated into a 1:1 ratio in both populations (Table 4.7) fitting a single locus control model. The segregation of hilum colour also fitted the expected 1:1 ratio in the LR-26 population (Table 4.7). However, in LR-59 (Table 4.7), the ratio between the two types was highly distorted as the dark hilum colour (Lcu type) was more predominant than the light hilum colour (Ler type). A two-loci model on hilum colour was further tested. The test results (Table 4.8) showed that in LR-59, the segregation of hilum colour could fit a 3:1 segregation; while in LR-26, this two-loci model hypothesis was rejected.

Table 4.5 Estimates of variance components and broad sense heritability for three quantitative seed quality traits in the LR-26 RILs (Eston x IG 72815) based on results from four environments in Saskatchewan.

Variance component	Thousand seed weight	Sucrose£	TRFO£
$\sigma^2 e$	1.38	0.002	0.004
$\sigma^2 g$	20.31	0.03	0.22
$\sigma^2 g*e$	2.15	0.01	0.21
$\sigma^2 p$	21.08	0.04	0.31
H^2	0.96	0.89	0.72

$\sigma^2 e$: environmental contribution to total phenotypic variation; $\sigma^2 g$: genetic variance; $\sigma^2 p$: total phenotypic variation; H^2: broad sense heritability; TRFO: Total raffinose family oligosaccharide; £: Sucrose and TRFO were tested in three environments

Table 4.6 Estimates of variance components and broad sense heritability for three quantitative seed quality traits in the LR-59 RILs (Eston x L01-827A) based on results from four environments in Saskatchewan.

Variance component	Thousand seed weight	Sucrose£	TRFO£
$\sigma^2 e$	1.37	0.02	0.42
$\sigma^2 g$	13.61	0.07	0.08
$\sigma^2 g*e$	0.15	0.07	0.25
$\sigma^2 p$	13.85	0.10	0.18
H^2	0.98	0.69	0.42

$\sigma^2 e$: environmental contribution to total phenotypic variation; $\sigma^2 g$: genetic variance; $\sigma^2 p$: total phenotypic variation; H^2: broad sense heritability; TRFO: Total raffinose family oligosaccharide; £: Sucrose and TRFO were tested in three environments

Table 4.7 Chi-square test results for segregation ratios of monolocus control for seed quality traits among RILs of LR-26 and LR-59.

	Hilum color of LR-26		Cotyledon color of LR-26		Hilum color of LR-59		Cotyledon color of LR-59	
	Darker	Lighter	Yellow	Red	Darker	Lighter	Yellow	Red
Source	Lcu	Ler	Lcu	Ler	Lcu	Ler	Lcu	Ler
Expected	84	84	86	86	33	33	33	33
Observed	78	88	78	94	56	10	34	32
X^2	0.60		1.49		32.1		0.06	
Probability	0.44		0.22		<0.0001		0.81	

X^2: Chi square test value. Both traits were tested at 1:1 ratio, all tests had a degree of freedom of 1

Table 4.8 Chi-square test results for segregation ratios of a two-loci control model on hilum colour among RILs of LR-26 and LR-59.

	Hilum color of LR-26		Hilum color of LR-59	
	Darker	Lighter	Darker	Lighter
Source	Lcu	Lcu	Lcu	Ler
Expected	110.7	55.3	59.33	29.67
Observed	78	88	56	10
X^2	13.93		0.56	
Probability	0.01		0.50	

X^2: Chi square test value. Segregation on pod dehiscence was tested at 3:1 ratio, tests were under a degree of freedom of 1

4.4 Discussion

Lentil plays a role in global food security. It is a staple source of vegetative protein and a traditional food source in many parts of the world. Therefore, the improvement of seed quality traits is a desirable goal in lentil breeding. CWR species have great potential in seed/grain and fruit quality improvements with a diverse compositional variation among various crop species as reviewed in Fernie et al. (2006), however, the use of CWR in the improvement of such complex traits remain underexploited (Hajjar and Hodgkin, 2007). This chapter presents the results of phenotyping for visual seed quality and seed nutritional composition in a population derived from a cross between a domesticated lentil (Lcu) and a tertiary genepool species of Ler.

In lentil, seed size is arguably the most important criteria for market classification and is a factor in cooking efficiency (Bhatty, 1984; Erskine et al., 1985). Seed size is considered a highly heritable polygenic trait in lentil (Abbo et al., 1991). High heritability estimates for seed weight have been repeatedly reported in cultivated lentil (Bicer et al., 2004; Erskine et al., 1985), as well as interspecific combinations from Lcu x Ler (Abbo et al., 1991; Singh et al., 2017), and Lcu x its progenitor species *L. orientalis* (Lor) (Abbo et al., 1991). As the market preference for a particular seed size depends on the original type in a particular market (Muelhbauer et al., 2009), selection for seed size is important for a lentil exporting country with multiple target markets.

In both LR-26 and LR-59, the interspecific RILs were derived from a medium sized Lcu (TSW of Eston was over 30 g) x small sized Ler (TSW was 5-6 g for IG 72815 and L01-827A) combination; and TSW segregated in an additive manner within this diverse range in both populations. A high broad-sense heritability estimate for TSW (0. 96 and 0.98, respectively) was found in both populations (Table 4.5, Table 4.6). Despite the strong genetic control over this trait among the RILs, a highly significant GxE interaction was still found in this study. As good soil fertility, particularly with regard

to carbon and nitrogen, and sufficient moisture are required for seed filling in grain legumes (Salon et al., 2001), this GxE interaction suggests that the RILs are responsive to the environmental factors in a given site/year but the high heritability indicates it is predictable.

Carbohydrates are the major constituents of a lentil seed, accounting for more than 50% of seed mass (Costa et al., 2006). After starch, soluble carbohydrates such as sucrose and RFO are the main components and they have an effect on dietary quality. Both sucrose and RFO play a physiological role as storage sugars in seed and are deposited during the maturation stage. They also act as transportation sugars between tissues. Sucrose is an important source of sweetness and energy from lentil consumption. While high levels of RFO consumption have been considered a bloating factor because of indigestibility (Fleming, 1982), RFO in lentil seeds are also considered positive as a type of prebiotic dietary fibre (Agil et al., 2013; Dwivedi et al., 2014; Johnson et al., 2013). Overall, the concentration of these two major soluble sugars are important indicators of carbohydrate quality of many pulse crops (Wang et al., 2003), including lentil.

Tahir et al. (2011b) reported a positive correlation between seed RFO level and sucrose concentration in lentil, as RFOs are derivatives of sucrose (Peterbauer and Richter, 2001). Tahir et al. (2011b) conducted a series of field trials in Western Canada screening a collection of global elite lentil (Lcu) varieties and landraces and reported that in lentil (Lcu) seeds, sucrose concentration ranged from 0.7 to 2.4 g/100g flour and TRFO concentration ranged from 4.6 to 6.6 mmole/100g flour (Tahir et al., 2011b). In a related study (Tahir et al., 2012), Ler was reported to have almost half the TRFO concentration of lentil cultivars, the lowest level among a diverse collection of Lens species. In this study, among the individual RILs of LR-26 and LR-59, novel variations with lower concentrations of sucrose (Figures 4.4 and 4.5) and TRFO (Figures 4.6 and 4.7) were observed in both populations, compared to the existing selection pool of

cultivated lentil (Tahir et al., 2011b). This result suggests that hybridization with Ler may be used to introduce new variation in the level of these two important soluble carbohydrates to the current cultivar lentil genetic background. This novel variation from introgression allows for selection for a lower level of TRO lentil should this be desired. A very similar heritability estimate for sucrose (0.89) was reported among lentil cultivars in Tahir et al. (2011a); while the broad sense heritability estimate for TRFO in the interspecific RILs (0.85) was a bit lower (0.72 and 0.42 from LR-26 and LR-59, respectively) than that reported within lentil cultivars (0.85; Tahir et al., 2011a). In general, these seed quality traits had a higher broad sense heritability (Table 4.5 and 4.6) compared to the agronomic traits (see Chapter 3, Table 3.8 and 3.9). The results suggest higher stability from the seed quality traits and a potential to harness genomic-assisted selection.

The genotypic degrees of freedom for sucrose and TRFO was lower than for TSW in both populations due to the limited numbers of seeds harvested from the RILs and the two accessions of wild parents. Previously, it was pointed out that the low seed set in LR-59 may be attributed to possible incompatibility issues which led the low viability and fertility of interspecific progeny (Fiala et al., 2009; Vail and Vandenberg, 2011). In addition to that, even though all the RILs were bagged before the pod-setting stage, the seed-loss from pod dropping, pod dehiscence and pod shattering may still have led to some level of reduced yield in the populations.

In this study, TRFO was estimated from the bulk content of all galactosyl-sucrose oligosaccharides in seeds. Tahir et al. (2012) found that other than having the lowest TRFO content among *Lens* spp., a different proportion among three major members of RFO (raffinose, stachyose and verbascose) with a higher verbascose and lower stachyose content could also be found in Ler genotypes. The results suggest a different genetic control underlying the biosynthesis of RFOs between Lcu and Ler. Their results were generated using a high-performance size exclusion chromatography method

which was beyond the scope of this current project. However, further assessment on the composition of RFO in the interspecific RILs may help elucidate the different genetic pathway between the two parental species. Furthermore, an improved high-performance anion exchange chromatography with pulsed amperometric detection method has been reported to be more accurate in characterizing the composition of RFO in chickpea (Gangola et al., 2014) and may be further applied in lentil to allow for better resolution to understand the inheritance behind the segregation.

Two visual qualitative seed quality traits, cotyledon colour and hilum colour, were characterized in these RILs. Cotyledon colour is a major characteristic dictating lentil market value and classification in the global market. Three cotyledon colours: red, yellow and green, are found in lentils. The red cotyledon type is the most common, composing roughly 80% of the total global lentil market. Although a three-gene model was proposed for genetic control of these three types of cotyledon colour (Sharma and Emami, 2002), the cotyledon colour in lentil is generally accepted to be under the control of two genes (Wilson et al., 1970; Slinkard, 1978). That is, the red vs yellow cotyledon colour is under genetic control of the *Yc* locus, where the allele *yc* produces yellow cotyledons and the dominant *Yc* allele gives red cotyledons. The green cotyledon colour is regulated by a second, recessive inhibitor gene, namely *i-yc*, while *I-yc* has no inhibitory function (Slinkard, 1978). Wild lentils come in both red and yellow cotyledon types and in the two populations examined here, the wild parents were red while the cultivated parent was yellow. There were no green cotyledon lines in either of the populations examined as both parents were fixed for the *I-yc* allele. The segregation of red and yellow cotyledon colour fitted a single-locus control model in both populations (Table 4.7). In a separate study, Singh et al. (2014b) also concluded there was a single dominant allele controlling red cotyledon colour based on their Lcu x Ler F_2 progeny.

The colour around the hilum scar segregated into a darker and visually self-coloured type (Lcu type) or lighter-coloured type (Ler type) among the RILs of LR-26 and LR-59. The hilum on dicot seeds is where the seed was attached to the funiculus, and the hilum colour is an important factor in classification and identification for many legumes species (Gasim et al., 2004). The pigmentation of the hilum may affect market preference of some edible legumes. For example, in soybeans, the light-hilum type varieties are much preferred for the human consumption market over the brown-hilum type (deMan et al., 1987), especially for white soy products such as tofu (bean curd) and soymilk. In lentil, the darker-colour hilum type (Lcu type) is generally the same colour as the seed testa.

While the segregation of these two types of hilum colour in LR-26 fitted a single-locus control model (Table 4.7); in LR-59, the population was highly distorted towards the Lcu type, darker colour (Table 4.7). Pod dehiscence was also distorted towards the Lcu type in LR-59 (Table 4.7). A moderately positive correlation was found between pod dehiscence and hilum colour in both LR-26 (phi coefficient=0.4 between pod dehiscence and hilum colour), and LR-59 (phi coefficient=0.45 between pod dehiscence and hilum colour). Thus, the distortion of hilum colour in the smaller population LR-59 may be associated with (un)conscious selection for non-dehiscent types during population development, resulted in higher chance of the darker hilum colour (Lcu type) being retained in LR-59.

Overall, to fully benefit from the introgression of fragments of a wild genome for lentil improvement, further exploration of the genetic components at a genomic level and the application of genomic-assisted pre-breeding are critical. For a more efficient utilization of the interspecific breeding materials, the phenotyping data collected in this chapter will be combined with genotyping results (Chapter 5) to help better understand the genetic regions controlling these phenotype variations (Chapter 6). To develop more interspecific populations and introgression maps would be major tasks for lentil

breeders for more efficient use of wild introgression. By increasing the diversity of interspecific combination, breeders would have a better chance to find desirable yet easily manageable breeding materials. Lastly, variation in additional seed quality traits may further be assessed among these and other diverse CWR and interspecific-derived hybrids. Evaluating seed nutrition quality traits of importance in lentil, such as protein content, protein quality and micronutrient composition, among diverse interspecific progeny, and could help identify the other valuable sources of novel variability for lentil improvement.

Prologue to Chapter 5

Generally, the hybridization between two species has brought additional genetic variation into the offspring for these tested quantitative agronomic and seed quality traits and most of the traits were detected with at least a significant ($p<0.01$) genotypic effect at certain single environment (Tables 3.3, 3.4, 4.3 and 4.4) with few exceptions such as DTE at CSSF15, DTF at CSSF15 and RP at CSSF14. The results suggest that the phenotyping data (presented in Chapter 3 and Chapter 4) can be used in combination with genotyping results of the interspecific RILs.

To help unravel the impacts of Lcu x Ler introgression at the genomic level, and to construct a linkage map for further marker-trait association analysis, RILs of LR-26 were genotyped as reported in the next chapter. The resulting genotypic data and linkage map of Chapter 5, in combination with the phenotypic data collected in Chapters 3 and Chapters 4, will be used for quantitative trait loci (QTL) mapping in Chapter 6.

CHAPTER 5 CONSTRUCTION OF AN INTROGRESSION GENETIC LINKAGE MAP OF LR-26 USING GENOME-WIDE HIGH-DENSITY SINGLE NUCLEOTIDE POLYMORPHISM VARIANTS

5.1 Introduction

To help bring desired disease resistance from the wild into cultivated lentil, the Pulse Crop Research Group at the University of Saskatchewan developed two Lcu x Ler derived RIL populations: LR-26 and LR-59 (Vail and Vandenberg, 2011; Vail et al., 2012; Fiala et al., 2009; Podder, 2012). From previous phenotype evaluations in both agronomic (Chapter 3) and seed quality (Chapter 4) characteristics, the potential benefits of the expanded genetic base have been assessed, while some possible issues such as segregation distortion, heterogeneity, genome incompatibility and linkage drag were also found within the populations.

Through genotyping and linkage analysis of these interspecific RILs, it would be possible to further investigate the genome level of organization resulting from CWI. For example, Zamir and Tadmor (1986) reported genetic segregation distortion in several intraspecific and interspecific *Lens* spp. F_2 populations using only a handful of isozyme and morphological markers. By using SNP markers, the most abundant type of genetic variations in nature, a high-density, genome-wide linkage map can be generated, providing better resolution of the genomic composition of the interspecific RILs.

Linkage maps are constructed by estimating the recombination ratio during development of a genetic population and can be used to calculate the relative position of polymorphic markers on the genome (Collard et al., 2005). The objective of the research reported in this chapter was to genotype a Lcu x Ler derived interspecific RIL population and to construct a high-density genetic linkage map using GBS-derived genome-wide SNP markers.

5.2 Materials and methods

5.2.1 Genotyping and SNP discovery of LR-26 using a two-enzyme GBS method

Genotype-by-sequencing (GBS) is a genome-wide sequencing approach to SNP variant discovery (Elshire et al., 2011), and it has wide application in plant genetics and breeding. In this study, the two-enzymatic GBS method developed by Poland et al. (2012) was applied to LR-26 to reduce the data complexity while still maintain genome-wide coverage of the SNPs.

5.2.1.1 Plant materials

Samples of genomic DNA were extracted from approximately 50 mg of fresh leaf tissue from a single plant of each RIL of LR-26. By the time of the trial, the population of LR-26 has gone through more than 10 generations of self-pollination. Tissue was collected from the field at STH15 during the vegetative growth stage and placed in labeled 5 ml Eppendorf tubes. In cases where there was not enough tissue available from a single plant, tissue was bulked from multiple plants to reach to the amount needed. For those lines that had not emerged in the field, tissue was collected from single plants grown in a mixture of 1:1 Sunshine Mix #3 and Sunshine Mix #4 (Sun Gro Horticulture Ltd., Canada) in a growth cabinet.

5.2.1.2 GBS library preparation

Tissue samples collected from each RIL of LR-26 were freeze-dried, and genomic DNA was extracted using a DNeasy® 96 Plant Kit (Cat# 6918, Qiagen, Germany) following the manufacturer's instructions. DNA quality was checked on a 1% agarose gel and DNA quantifications were done using a Quant- iT PicoGreen dsDNA assay kit (Life Technologies, USA) on a FLUOstar Omega fluorometer (BMG Labtech, USA). The preparation of the two-enzyme GBS sequencing library was done following the *Pst*I-*Msp*I GBS protocol of Poland et al. (2012) following the steps below:

Normalization of DNA: All DNA samples were normalized to 20 ng/μl after quantification as a working concentration. For each sample, 10 μl of genomic DNA working solution was used for a total of 200 ng of DNA.

Restriction of DNA: For each sample, the restriction reaction contained 8 units of *Pst*I-HF, 8 units of *Msp*I and 10X NEB buffer 4 (all from New England Biolabs, USA), and the final volume of the restriction digest was adjusted to 20 μl using autoclaved Milli-Q® Ultrapure-water (Milli-Q® Reference Water Purification System, EMD Millipore Corporation). Digestion was done using a thermocycler. The reaction was incubated at 37°C for 2 hr, followed by 65°C for 20 min to denature the restriction enzymes before holding at 8°C.

Ligation to GBS adapters: A ligation reaction was done within the same tube as the restriction reaction. A ligation mastermix was prepared first, containing 1 mM ATP and 200 unit of T4 DNA ligase and 1X NEB buffer 4 (all from New England Biolabs, USA). The adapter mix was prepared in a 96-well plate. Each well of the adapter mix contained 0.02 μM of barcoded adapter (Adapter 1 with *Pst*I cutting site) and 3 μM of common adapters (Adapters 2 with *Msp*I cutting site). The barcodes and adapter sequences were designed and optimized for Illumina pair-end sequencing (Illumina, USA). The oligonucleotide sequences of the Illumina adapters used for their library prep kits are available on the Illumina client support website at https://support.illumina.com/downloads/illumina-customer-sequence-letter.

Into each ligation reaction tube, 15 μl master mix and 5 μl of pre-mixed adapter mix were added to the 20 μl of restriction digest product. At this step, the final volume of each reaction was 40 μl. The ligation mixture was incubated at 22°C for 2 h, followed by 20 min at 65°C to inactivate the ligase, using a thermocycler.

Multiplexing of samples: There were 43 genotypes pooled in each of the four sequencing libraries. For each sample, 10 μl of adapter-ligated DNA was taken from the ligation mixtures. For each library, adapter-ligated DNA from each of 43 genotypes was

pooled in a tube, and the final volume was adjusted to 500 µl with autoclaved ultrapure water. Four hundred µl of bulked DNA sample was taken out of the 500 µl of each library to be purified using a QIAquick® PCR purification kit (Qiagen, Germany) following the manufacturer's instructions. After this step, approximately 120 µl of purified DNA from each library was used for further processing.

Library amplification: This step was to increase the number of fragments in the range of 200 bp to 500 bp in the sequencing library, which is the ideal size range for the Illumina® sequencing platform. The primer sets were designed according to the sequences of Adapters 1 and 2. So, only the fragments of this size range that contained both the *Mst*I and *Pst*I restriction sites would be amplified.

A mastermix for PCR amplification was prepared. For each library, 8 PCR reactions were made, and each reaction contained 10 µl of purified pooled DNA, 1X NEB MasterMix, 0.8 µM of Illumina-PE® primer 1 and 0.8 µM Illumina-PE® primer 2 (Illumina, USA). The final volume of each PCR reaction was adjusted to 25 µl with autoclaved ultrapure nano water. PCR amplification was done using a thermocycler under with the following program: denaturation at 95°C for 30 sec; 16 cycles of 30 sec at 95°C, 20 sec at 62°C, 30 sec at 68°C and 5 min at 72°C, and a final hold at 8°C, using a thermocycler. All 8 PCR reactions from one library were pooled and purified using a QIAquick PCR purification kit (Qiagen, Germany), and the purified DNA was eluted in 30 µl of the elution buffer.

Quality control of the library: This step was to check the size range of DNA fragments of the library using capillary electrophoresis. The quality check of each library was performed using the Agilent 2100 Bioanalyzer® system and an Agilent DNA 1000 Kit (Agilent Technology Inc., USA) following the manufacturer's instructions. A successfully prepared library should contain DNA fragments within the suitable range of 200 bp to 500 bp without much adapter-dimer, and DNA concentration should exceed 3 ng/µl to meet the sequencing requirements.

5.2.1.3 Next-generation sequencing

The prepared sequencing libraries were sent to the sequencing facility of McGill University's Génome Québec Innovation Centre following the sample submission requirements in http://www.gqinnovationcenter.com/index.aspx/servise. Sequencing was conducted on an Illumina HiSeq platform (Illumina, USA). The sequencing results were provided by the sequencing facility in a FASTQ format file.

5.2.1.4 Sequencing output analysis

The GBS results of four sequencing libraries constructed from 172 RILs of LR-26 (each genotype had two sequencing replications); and the GBS results of two parents of LR-26 (IG 72815 and Eston) were included in the SNP calling process. A quality check of the raw sequencing files was done before processing using the FASTQC program (http://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc). This step was to generate a statistical report of the sequence reads and quick overview to detect potential problem areas of the raw sequences.

The GBS data analysis was done following an in-house GBS pipeline (http://carolyncaron.github.io/GBSpipeline/) (Wong et al., 2015). This pipeline was developed by optimization of commonly used open-source programs for use in lentil genome research. This pipeline runs from the command line using Perl scripts as follows:

De-multiplexing: By aligning the Illumina barcode adapters, this step counts the number of all the pair-end sequence reads from GBS, and de-multiplexes all the reads of each sample into separate files. The sequences of the Illumina barcodes and the PstI and MspI restriction sites were required for this step.

Trimming of reads: This step was to remove the sequence of the adapters from the pair-end sequence reads by running the Trimmomatic program (http://www.usadellab.org/cms/?page=trimmomatic). Only those reads with a minimum size of 36 bp after trimmed were retained at this step.

Alignment of reads: All sequencing reads after trimming were aligned to physical positions in the lentil reference genome version 1.2 (cultivar CDC Redberry) (http://knowpulse.usask.ca/portal/lentil-genome) using the Bowtie2 program (http://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.2.3/). Due to low number of reads mapped to the reference genome at this step, three RILs, LR-26-83, LR-26-115 and LR-26-152, were removed from the list of genotypes for further processing and analyses.

SNP calling: The final step was the SNP discovery for each sample using SAMtools and BCFtools; whereby SAMtools collected summary information and BCFtools made the variant calls. Both programs can be downloaded at: http://www.htslib.org/download/.

5.2.2 Processing SNP markers

SNP marker filtering: The SNP matrix determined from the GBS genotyping of 169 RILs (see reads alignment step of GBS analysis section) was used for the construction of a genetic linkage map. To assure the quality of the SNP dataset, four specified parameters were first used for SNP filtering. First, a minimum read depth of five was required for SNP calling, violating sequences were treated as missing. Second, SNP markers with over 25 % of the genotypes (more than 42 RILs) missing were removed. Third, SNP markers which were monomorphic for the parental alleles were removed. Last, highly distorted SNP markers with minor alleles only found in fewer than 10 RILs were removed.

After filtering, the genotyping output was set in an ABH format, where the "A" represented SNPs called from Eston (Lcu) genome and the "B" represented SNPs called from IG 72815 (Ler) genome. The "H" represented heterozygous calls, where SNPs with a read depth greater than 10 and with both parental alleles present were considered heterozygous.

The distribution of heterozygous calls among genotypes was further examined, and two genotypes, LR-26-113 and LR-26-312, were removed from the list of genotypes at this step due to extremely high genome-wide heterozygous coverage. The remaining 167 lines were used for linkage map construction (a list of genotypes can be found in Appendix C). LR-26 had gone through more than 10 generations of self-pollination to reach the theoretical inbred state, therefore, the heterozygous markers were not recognized for linkage analysis by most of the mapping algorithm. Therefore, the heterozygous SNPs were treated as missing data to transform the ABH format into homozygous AB format for initial map development.

Nomenclature of the SNP markers: The nomenclature of the SNP markers was based on their physical position on lentil reference genome, version 1.2 (http:/knowpulse.usask.ca /portal/jbrowse/Lentil). The number before the capital letter "C" of each SNP marker represents the corresponding chromosome on the lentil genome, and the number following the capital letter "C" indicates the physical position in base pairs (bp). SNPs with a name starting with the letters "Ctg" were aligned to unmapped contigs of the lentil reference genome, version 1.2, and the number after "Ctg" was based on the LcContig number (Lc stands for *Lens culinaris*).

5.2.3 Marker distribution visualization

A JavaScript based chromosome visualization tool (CViTjs) (Cannon and Cannon, 2011) was used to view the distribution of marker frequency, including the minor alleles, based on the lentil genome v1.2 (source: https://github.com/LegumeFederation/cvitjs). An in-house script developed by the Pulse Crop Bioinformatics group at the University of Saskatchewan by optimization of commonly used open-source programs for use in lentil genome research was used. After filtering out the low-quality SNPs calls, but retaining all the minor alleles (see section 5.2.2), 5,392 SNPs were harvested for generating the plot. Further, after removal

of the minor alleles, 5,293 SNPs of 169 RILs were plotted using CViTjs to show the location of heterozygous markers.

5.2.4 Linkage analysis

After plotting for the marker distribution plots, the heterozygous SNPs were treated as missing, and all the SNPs were filtered again to remove SNPs with too many missing data points. Two individual genotypes, LR-26-113 and LR-26-312, were removed from linkage analysis at this step due to genome-wide heterozygosity resulting in too many missing datapoints for mapping. Eventually, 4,726 SNPs were retained, and the genotyping information was entered into QTL IciMapping version 3.2 (Meng et al., 2015) for linkage map construction. Bin functionality of IciMapping version 3.2 was executed first to reduce redundancy of the high-density markers. Redundant markers were those with a correlation coefficient of 1; these markers were completely linked and did not provide extra information for linkage mapping. The 4,726 markers segregated into 1,076 bins and one representative marker from each bin was kept for map construction (Appendix D & E).

The threshold logarithm of odds (LOD, the log of likelihood ratio) was set at 24.5. Different map algorithms were tested for marker ordering, and the Kosambi map function (Kosambi, 1944) was used to calculate marker distance. The two-opt algorithm ("nnTwoOpt" function in the software), which is efficient in solving the "traveling salesman" problem (Lin and Kernighan, 1973; Laporte, 1992), was chosen as it provided the shortest map distance output. A total number of 1,071 unique markers were mapped. Map output was generated using MapChart 2.30 software (Voorrips, 2002).

5.3 Results

5.3.1 Distribution of the Lcu and Ler genome fragments

Before removing the highly distorted minor alleles, a total of 5,392 SNPs were harvested from the first three filtering criteria as described in section 5.2.2. These SNPs

mapped across all 7 Lcu chromosomes, with very few missing regions, based on their physical positions on the lentil reference genome (version 1.2) (Figure 5.1). From the results of allele segregation, most chromosomes have a nearly equal segregation ratio between two parental species (a ratio of 40 to 60 % of Lcu allele at each SNP) (Figure 5.1). In some regions, however, especially on chromosomes 1, 3 and 5, Lcu alleles seem to have been favored. Distorted distribution towards Ler alleles appeared to cluster on the tails of chromosomes 4, 5 and 6 (Figure 5.1). After filtering out the SNPs where the minor allele frequency was < 6 %, the remaining 5,293 SNPs were retained for further analysis. The distribution of high confidence heterozygous loci (both alleles represented in >10 calls at a locus) was concentrated on chromosomes 1, 5 and 7 (Figure 5.2).

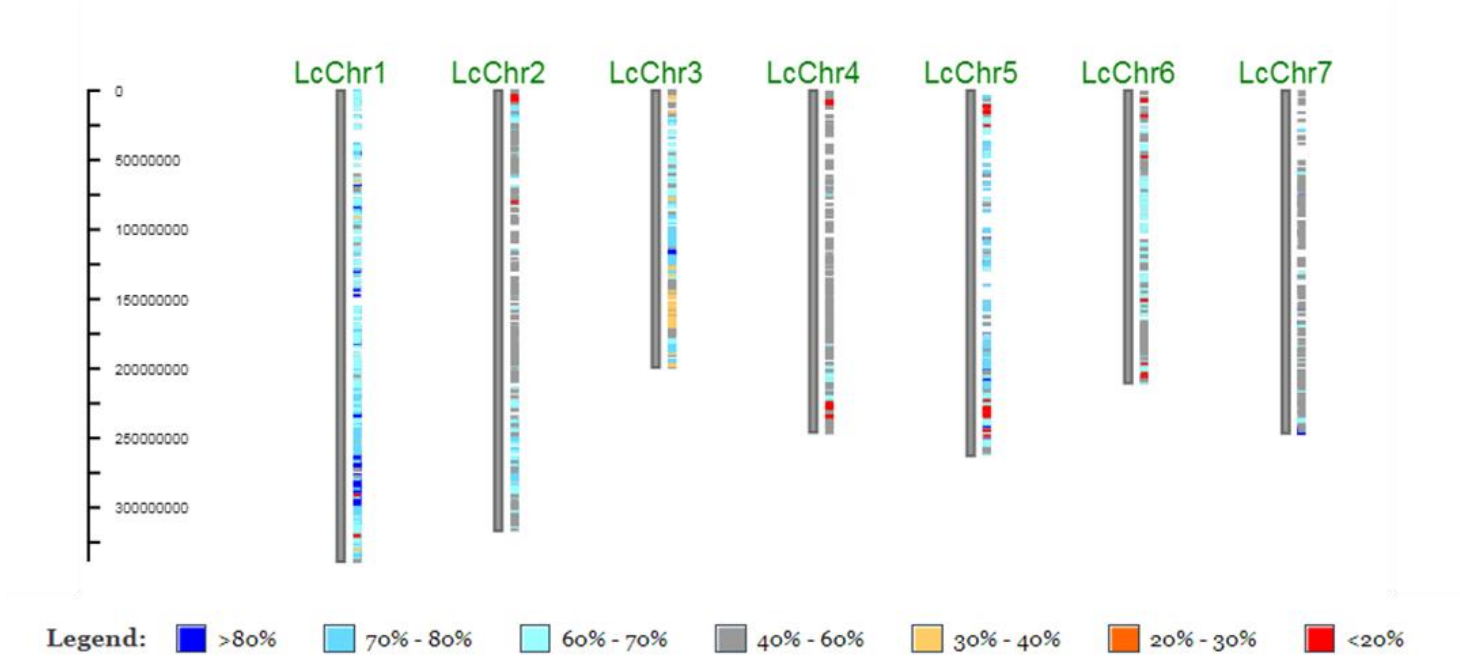


Figure 5.1 SNP marker distribution and allele segregation in RIL population LR-26 (derived from Eston x IG 72815) based on physical positions across the lentil reference genome assembly (version 1.2). The coloured bars represent loci in the genetic linkage map and the colors indicate the proportion of the Lcu genome. Number on the y-axis represents the physical position in base pairs for each chromosome.

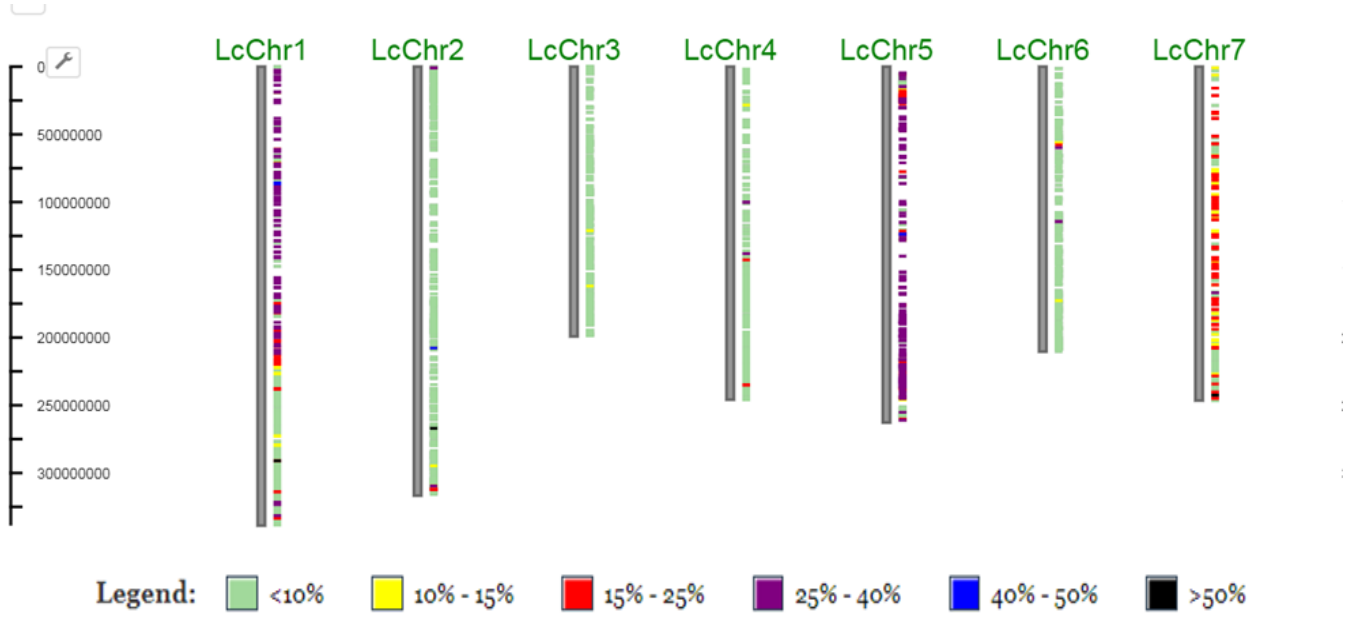


Figure 5.2 Distribution of heterozygous SNP loci within the LR-26 population (derived from Eston x IG 72815), based on both alleles being represented in >10 calls of the raw read data, across the lentil genome assembly (version 1.2). The colored bars represent loci and the colors indicate the proportion of heterozygotes across the RIL population. Numbers on the y-axis represents the physical position in base pairs for each chromosome.

5.3.2 Developing a linkage map based on the recombination of SNPs in RIL population LR-26

An interspecific linkage map, comprised of 1,071 unique SNP marker bins with a total length of 984.26 cM, was generated for LR-26 (Table 5.1; Figure 5.3). High-density bins segregated into seven linkage groups (LG) with an average distance of 0.92 cM between markers (Table 5.1). A list of markers in each bin can be found in Appendix E, and the unique SNP markers (bins) and positions can be found in Appendix D. There was one very large linkage group (LG1) that contained SNP markers representing regions of Lcu chromosomes 1 and 5 and all of 7 based on the lentil reference genome (Table 5.1; Figure 5.3) and could not be logically separated further. The rest of the SNPs representing Lcu chromosomes 1 and 5 were grouped into single linkage groups (LG 2 and LG 6, respectively). The Lcu alleles showed slightly higher overall frequency across the whole map (Table 5.1).

Table 5.1 High-density linkage map of LR-26 using SNPs discovered using GBS.

Linkage group (LG)	Corresponding Lcu chromosome€	Number of bins (loci)	Length (cM)	Average marker interval (cM)	% of Lcu per SNP locus	% of Ler per SNP locus	% of missing calls per SNP locus
1	1,5,7	262	252.22	0.96	68%	28%	4%
2	1	19*	22.73	1.20	60%	32%	8%
3	2	193	153.23	0.79	61%	34%	5%
4	3	208	168.3	0.81	65%	34%	1%
5	4	166	161.05	0.97	55%	43%	2%
6	5	77*	102.16	1.33	62%	31%	7%
7	6	146	124.57	0.85	59%	39%	2%
Whole genome		1071	984.26	0.92	62%	35%	3%

€: Based on physical marker position of the lentil reference genome assembly version 1.2.
*: The rest of this chromosome is found in LG1.

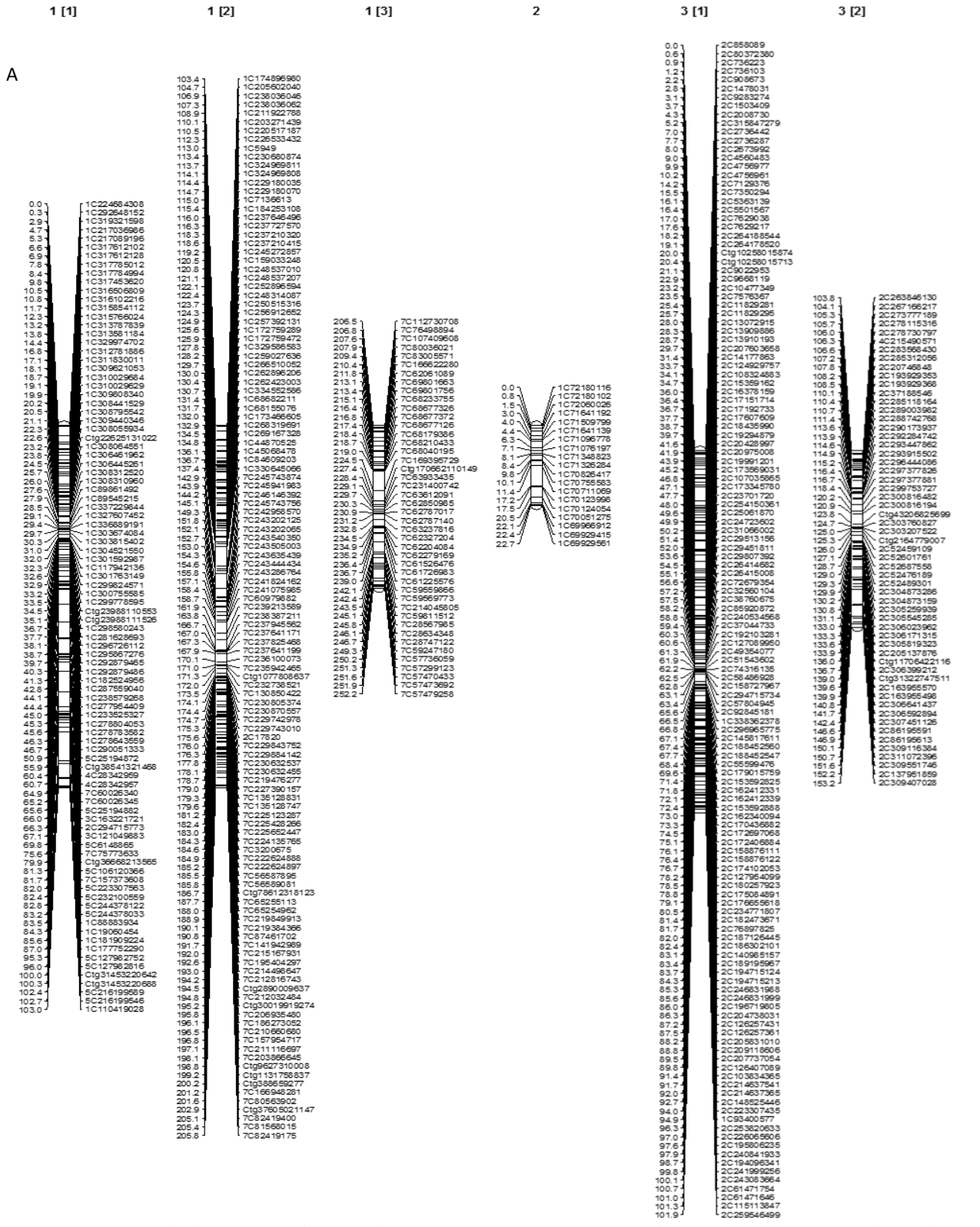
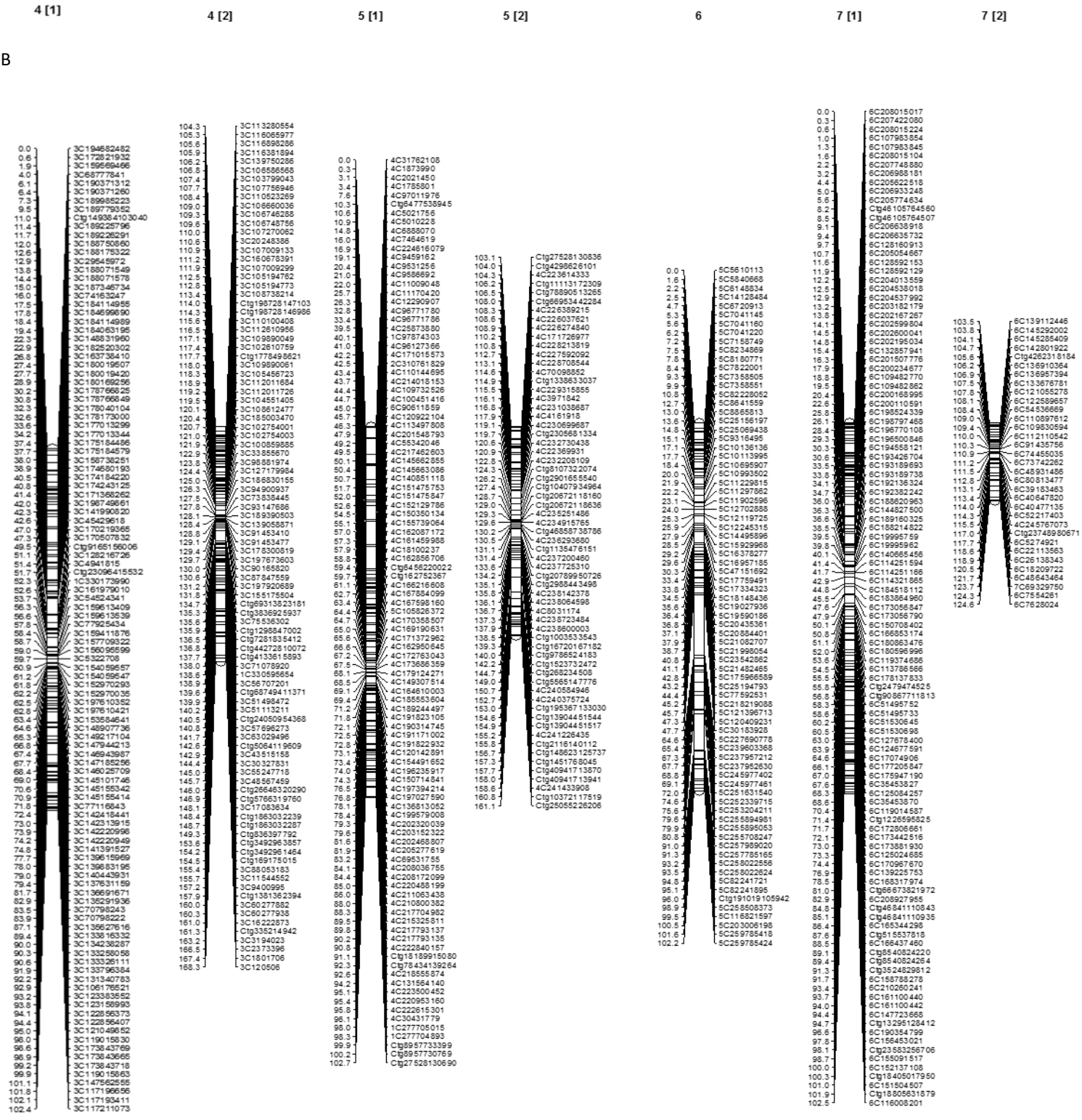


Figure 5.3 Genetic linkage map of *Lens culinaris* (variety Eston) x *L. ervoides* (assession IG 72815) derived recombinant inbred lines population LR-26 based on SNPs generated using GBS. The number on top of each linkage group corresponds to the number indicated in Table 5.1. Panel A: Linkage groups 1-3; Panel B: Linkage groups 4-7. Number inside the bracket indicated the order within the linkage group.

B



5.4 Discussion

The hybridization of Lcu and Ler to create RIL populations has brought desirable disease resistance to the cultivated lentil genetic base (Fiala et al., 2009; Podder et al., 2012; Singh et al., 2017; Vail et al., 2012), and created an abundance of other phenotypic variation (Tullu et al., 2013; Singh et al., 2013; Chapters 3 and 4 of this thesis). Although the crossing barriers can be overcome in Lcu x Ler crosses through embryo rescue (Fratini and Ruiz, 2010), challenges such as low fertility and viability (Fiala et al., 2009), and phenotypic segregation distortion (Tullu et al., 2013; Chapters 3 and 4 of this thesis) still exist. Genomic analysis can help to clarify the genomic regions that are potentially problematic for breeders and better assess the potential use of CWI in lentil breeding. Thus, the objective of this chapter was to genotype the RILs of the mapping population LR-26, and to develop a genetic linkage map to explore the introgression patterns and that can be used in the analysis of genotype-phenotype associations (Chapter 6).

The viability among the interspecific lines is an important challenge in maintaining the population size of Lcu x Ler derived progeny, as was pointed out for another Lcu x Ler RIL population, LR-59 (Fiala et al., 2009). Ladizinsky et al. (1985) observed quadrivalent chromosome pairing during meiosis in the offspring from Lcu by Ler crosses, indicating abnormal homologous pairing in some chromosomes. It was concluded that a chromosomal interchange exists that distinguishes Lcu and Ler and is likely the major cause of the crossing barriers between the two species (Ladizinsky et al., 1985; Tadmor et al., 1987). Gujaria-Verma et al. (2014) identified a reciprocal translocation between chromosomes 1 and 5 of Lcu relative to Ler by indirectly comparing genetic linkage maps of an intraspecific Lcu RIL population (LR-18) and an intraspecific Ler RIL population (LR-66). Their study was based on the synteny of the markers with the *Medicago truncatula* genome. This translocation was confirmed in Bhadauria et al. (2017) using homologous sequences that allowed a direct comparison between Ler and Lcu. This reciprocal translocation would likely have caused a partially

pairing between the chromosomes of the two species resulting in quadrivalent formation during meiosis of the zygotes of Lcu x Ler hybrids (Ladizinsky et al., 1985). When a translocation heterozygote goes through meiosis, gametes derived from adjacent segregation are inviable; and result in lower viability. As a result, sterility issues including embryo abortion arise, increasing the difficulty to maintain the size of a Lcu x Ler population.

Mapping of the GBS reads onto the Lcu reference genome pointed to the possible existence of translocation heterozygotes in the form of a higher frequency of heterozygosity in the population on regions corresponding to Lcu reference genome chromosomes 1 and 5 (Figure 5.2). The development of LR-26 was through six generations of single seed descent and seeds from single plants were bulked harvest at the F_7 stage. They had gone through more than 10 generations of self-pollination at the time of the genotyping experiment without conscious selection. Lentil is a strict self-pollinated species (Wilson and Law, 1972) with cleistogamous flowers, therefore, the plants should have reached permanent homozygosity and the heterozygous calls would be unlikely to be associated with pollen contamination. Additionally, during the field observation in this study, some phenotypic segregations were still occurring in the supposedly fixed genotypes. One possible cause of such segregating phenotypes would be either residual heterogeneity due to early bulking or heterozygosity in the genotype. However, most DNA was derived from a single plant, and if from several, as heterozygosity should be detected randomly across the genome among the individuals of a heterogeneous population. The clustering of heterozygous SNPs on specific chromosomes suggests it is more likely due to the presence of segments of both parents in these regions with no recombination that impacted the RILs. In this case, the heterozygosity on Lcu reference genome chromosomes 1 and 5 are likely due to the existence of translocation heterozygotes.

The presence of tandem duplications or repeats in the genome is a third possible cause of heterozygous calls from GBS results. These repeats can end up collapsed in a genome assembly as they cannot be identified as separate segments. GBS reads are usually 100 to 300 bps in length, therefore, when a SNP variant is detected from a duplicated region, reads from an alternative read might also be mapped to this segment and result in a false call of a heterozygote. There is a large region of Lcu chromosome 7 that is suspected to be collapsed in the Lcu genome assembly because of a large duplication (Larissa Ramsay, personal communication). This is probably causing the increased number of heterozygous SNPs in the region of the LR-26 map that corresponds to Lcu chromosome 7 (Figure 5.2). These heterozygous SNPs were further treated as missing data for the QTL analysis, however, to meet the assumption of RIL population.

Additionally, while Mendel's segregation law describes the natural principle to ensure the parents contribute equally to the genetic background of their progeny, in the case of LR-26, there were distorted regions observed in several parts of the genome based on the SNP results (Figure 5.1). This was especially true in the case of regions corresponding to Lcu chromosomes 1, 3 and 5 where the markers skewed towards the Lcu alleles (Figure 5.1). In addition, at the ends of chromosomes 2, 4, 5 and 6, smaller regions of markers that were distorted toward Ler were detected. This violation of equal segregation is not unusual in interspecific progeny across plant species (e.g. Bliss et al., 2002; Ky et al., 2000; Zamir and Tadmor, 1986), and Eujayl et al., (1998) reported seeing significant marker distortion even in a cultivated lentil RIL population. In this study, since the differential adaptation of two species often leads to differential level of fitness of one genome over the other (Zamir and Tadmor, 1986), the distortion may hint at a role of environmental selection in altering allele frequency during the development of the population, resulting in a higher proportion of Lcu alleles in most parts of the genome. The segregation distortion towards Ler was only detected in smaller regions.

Since segregation distorters can be attributed to chromosomal meiotic drive (reviewed in Lyttle, 1991), further analysis of possible segregation distortion-related loci in these distorted regions may help clarify the distortion phenomenon.

In theory, genetic linkage groups should be composed of a group of physically linked polymorphic markers which segregated during chromosomal crossing over in meiosis during population development. However, in the linkage analysis results of LR-26 (Table 5.1; Figure 5.3), a group of SNPs that map to the lentil reference genome chromosomes 1, 5, and 7 could not be divided into independent linkage groups (Table 5.1; Figure 5.3). Rather than being physically linked, this genetic linkage was more likely to be attributed to the high number of heterozygous calls observed in these regions. These had been to be treated as missing data for mapping purposes which meant the markers could not be sufficiently separated by the mapping algorithm and resulting in one large linkage group (LG1) composed of pseudolinked markers across involving these three chromosomes. Other than the pseudolinkage in LG1, however, each other LG of LR-26 linkage map was grouped with SNP markers called from a single specific Lcu chromosome (Table 5.1).

The genetic linkage analysis of LR-26 allowed for a closer look at the genome level impact of the interspecies hybridization on the genome, and we further identified the challenges of developing permanent genetic populations. While breeders may put populations through many generations of self-pollination (in this case, more than seven generations) before bulking seed, the aberrant meiotic paring might continue to exist in the chromosome structure of the progeny. Besides, the abnormal homeologous pairing at mismatch regions due to the presence of chromosomal rearrangements such as reciprocal translocation, deletions, insertions and inversions often lead to a reduced frequency of meiotic crossovers (Martin et al., 2017). Therefore, linkage drag might become a hidden threat in the case when alleles for desirable traits are located in such regions alongside deleterious alleles.

In LR-26, for the selection of more accessible genotypes to be included in future use, breeders would look for lines that carry disease resistance with less wild introgression. However, if the loci associated with certain undesirable traits such as seed dormancy (larger DTE) or, pod dehiscence are located nearby the resistance genes in these two chromosomes 1 and 5, the chance to break the linkage would be lower.

Prologue to Chapter 6

A genome-wide and high-density interspecific genetic map of a Lcu x Ler RIL population using GBS derived SNP markers was generated and presented in Chapter 5. The map will enable researchers to further investigate the introgression regions which contribute to observed phenotypic variation. Despite the tremendous phenotypic variation observed in the progeny brought about by the hybridization between Lcu and Ler (Tullu et al., 2013; Singh et al., 2017, chapters 3 and 4 of this thesis), the map generated here helped reveal several challenges within the genomes of these RILs. The major challenges include segregation distortion of parental alleles, possible sterility/ lower viability, as well as the possible permanent level of heterozygosity within some of the RILs.

Mapping the genetic control of a specific trait can help determine which regions of the genome need to be tracked for a systematic introgression to minimize the linkage drag (Cameron et al., 2017). Therefore, in Chapter 6, results of QTL mapping in LR-26 will be presented for each segregating trait from chapters 3 and 4.

CHAPTER 6. TAGGING INTROGRESSION EVENTS ASSOCIATED WITH PHENOTYPES SEGREGATING AMONG LR-26 INDIVIDUALS

6.1 Introduction

Introgressiomics is a systematic approach for deliberate and optimal trait introgression using exotic materials by employing genomic and phenotypic data (Prohens et al., 2017). QTL mapping is a powerful tool to help track the introgressed regions underlying phenotypic variations in interspecific populations and enables breeders to manipulate this introgression-by-design approach.

LR-26 is an interspecific RIL population which was developed to introduce desirable and novel disease resistance genes from Ler, a CWR species from the tertiary genepool to cultivated lentil Lcu (Vail et al., 2012; Podder et al., 2012). The hybridization between these two genomes brought a lot more variability beyond this purpose and helped expand the genetic base of the current selection pool (Tullu et al., 2013; chapters 3 and 4 of this thesis). The consequences of the Lcu x Ler hybridization could be found in many additional quantitative traits of varying levels of importance, including phenology traits such as DTE, DTF, VP and RP, plant morphological trait such as PH, seed quality traits such as TSW, and seed compositional quality traits such as concentrations of TRFO and sucrose in seeds. According to these results (Table 3.4, Chapter 3 and Table 4.3, Chapter 4), variation in each trait can be aggregated to genotypic effects by environment, therefore, can be used for association analysis with the genotypes generated through GBS and the resulting genome-wide, high-density, SNP-based linkage map (Chapter 5).

The objective of this study was to identify the introgression regions associated with the phenotypic variation observed in Chapters 3 and 4 and better understand the genetics underlying these traits in an interspecific population. Since population size is a key factor in determining the detection power of QTL mapping, the larger population

size of LR-26 would enable a more precise QTL analysis than would be possible with the smaller LR-59 population. So, LR-26 was chosen as the mapping population.

6.2 Materials and methods

6.2.1 LR-26 phenotypic and genotypic data collection

The phenotypic data collected for agronomic traits, including DTE, DTF, VP, RP and PH, was described in Chapter 3 section 3.2; and for seed quality traits, including TSW, sucrose concentration and TRFO concentration, was as described in Chapter 4 section 4.2. The genotypic data was the SNP array that was used to generate the linkage map of LR-26 described in Chapter 5 section 5.2.

The segregation of monogenic traits was described in section 3.2.2 (flower colour and pod dehiscence) and section 4.2.2 (hilum colour and cotyledon colour). These data were used as morphological markers for linkage analysis with the genome-wide SNP variants detected from GBS results (Chapter 5) by coding the Lcu phenotype as 'A' and the Ler phenotype as 'B' and adding them to the LR-26 map using QTL IciMapping version 3.2 (Meng et al., 2015).

6.2.2 Phenotypic correlation coefficient test

A correlation analysis among phenotypes in a given environment was done in SAS software version 9.4 (SAS Inc., USA) using the Proc Corr statement and the Spearman's rank-order correlation coefficient method. Due to the disturbance to emergence time at CSSF15 from a cutworm infestation (Table 3.4 and Figure 3.4), the phenotypic data for DTE-CSSF15 were removed from the correlation coefficient test.

6.2.3 Genotype-phenotype association analysis

Files containing the linkage groups, genotypic data and phenotypic data can be downloaded from http://knowpulse.usask.ca/portal/LR-26-Marker-Bins-2018Mar15. The three files were entered into MapQTL version 6.0 (Kyazma B.V., Wageningen,

Netherlands) for QTL analysis. All parameters were maintained at the default setting and phenotypic scores were analyzed at each environment independently because of the significant GxE interactions (see sections 3.3.1 and 4.3.1). Input data used the average value of the three replications of each genotype at each environment.

For each phenotype, a simple interval mapping method was first used to estimate the logarithm of odd ratio (LOD) value of the association between each input phenotypic value and genome-wide SNP markers. The detection of LOD threshold values was done using permutation tests (1000 replication) in the genome-wide regions representing 95% confident intervals (Van Ooijen, 1999). The markers with highest LOD score at each QTL peak from the interval mapping test were chosen as cofactors in a multiple QTL mapping (MQM) model (Jansen et al., 1995) for final QTL detection.

6.3 Results

6.3.1 Phenotypic correlations

Correlations among all traits within each test environment are presented in Table 6.1. Significant correlations were found among most of the phenology traits (DTE, DTF, VP and RP) among most environments. Significant correlations were found consistently among the three seed quality characters across the different environments, as well as between the quality and phenology traits.

There was consistently a strong positive correlation between DTE and DTF among all the three tested environments. Negative correlations were found between DTE and VP in CSSF14 and STH15, while it was slightly positive in CSSF13 (Table 6.1). And since the calculation of VP was based on DTF and DTE, highly significant ($p<0.001$) positive correlations were found consistently between DTF and VP (Table 6.1) in all environments except in CSSF15 ($P<0.05$), where the emergence date was disturbed by a cutworm invasion (Table 6.1). The relationship between RP and the other phenology traits, however, was not consistent in this study (Table 6.1). At CSSF14, the correlation

Table 6.1 Spearman rank correlations for LR-26 RILs tested in four different site-years. Correlation coefficients and test estimation were made between two traits measured in same environment.

Trait	Environment	Days to flower	Vegetative Period	Reproductive Period	Plant Height	Thousand Seed Weight	Sucrose	TRFO
Days to Emergence	CSSF13	0.46***	0.02*	--	-0.32***	-0.50***	--	--
	CSSF14	0.55***	-0.61***	0.07ns	-0.40***	-0.57***	-0.20***	-0.11*
	CSSF15§	--	--	--	--	--	--	--
	STH15	0.59***	-0.11***	-0.01ns	-0.29***	-0.47***	-0.26***	-0.13***
Days to flower	CSSF13	--	0.85***	--	-0.19***	-0.47***	--	--
	CSSF14	--	0.25***	0.05ns	-0.22***	-0.47***	-0.21***	-0.10ns
	CSSF15	--	0.01*	-0.44***	-0.30***	-0.22***	-0.23***	-0.03ns
	STH15	--	0.70***	-0.29***	-0.15***	-0.54***	-0.33***	-0.31***
Vegetative Period	CSSF13	--	--	--	-0.04ns	-0.27***	--	--
	CSSF14	--	--	0.00ns	0.21***	0.23***	0.04ns	-0.001ns
	CSSF15	--	--	-0.06*	0.13ns	0.10ns	0.08 ns	-0.07 ns
	STH15	--	--	-0.32***	0.06 ns	-0.23***	-0.16*	-0.22**
Reproductive Period	CSSF14	--	--	--	-0.04ns	-0.10**	0.04ns	0.02ns
	CSSF15	--	--	--	0.19***	0.02 ns	0.04 ns	-0.05 ns
	STH15	--	--	--	0.07 *	0.03 ns	-0.07 ns	0.02 ns
Plant Height	CSSF13	--	--	--	--	0.40***	--	--
	CSSF14	--	--	--	--	0.36***	0.06ns	0.11ns
	CSSF15	--	--	--	--	0.34***	0.09 ns	-0.01 ns
	STH15	--	--	--	--	0.05 ns	-0.04 ns	0.00 ns
Thousand Seed Weight	CSSF14	--	--	--	--	--	0.39***	0.25***
	CSSF15	--	--	--	--	--	0.28***	0.33***
	STH15	--	--	--	--	--	0.39***	0.43***
Sucrose	CSSF14	--	--	--	--	--	--	0.45***
	CSSF15	--	--	--	--	--	--	0.43***
	STH15	--	--	--	--	--	--	0.17**

***: significant at P< 0.001, **: significant at P< 0.01, *: significant at P< 0.05, ns: not significant; TRFO: total raffinose family oligosaccharides; CSSF13: Crop Science seed farm at 2013; CSSF14: Crop Science seed farm at 2014; CSSF15: Crop Science seed farm at 2015; STH15: Sutherland experiment farm at 2015; §: DTE in CSSF15 was removed due to cutworm invasion occurred in this environment; £: Phenotypes of RP, sucrose and TRFO were not measured in this trial.

was not significant between RP and any other phenology trait. However, there was still a trend of a negative correlation between maturation time and flowering time, as RP was negatively correlated with DTF in CSSF15, and was negatively correlated to both DTF and VP in STH15 (Table 6.1).

Plant height (PH) was found to be negatively correlated with both DTE and DTF in all the environments where it was measured ($p<0.001$; Table 6.1). It was positively correlated to VP in CSSF14 and CSSF15; but a negative correlation was found between PH and VP in STH15, and the correlation was not significant in CSSF13 (Table 6.1). There was a significant positive correlation between RP and PH in CSSF15 and STH15; but not in CSSF14 (Table 6.1).

For the relationships among the seed quality traits measured, TSW had highly significant positive correlations ($p<0.001$) with both sucrose and TRFO (Table 6.1), and the relationship between sucrose and TRFO also showed highly significant positive correlations ($p<0.001$) across all environments tested (Table 6.1). A strong negative correlation existed between DTE and TSW, and between DTE and sucrose concentration. A weak negative correlation was found consistently between DTE and TRFO concentration among all the tested environments. A negative correlation was found consistently between DTF and the seed quality traits of TSW and sucrose concentration, while the highly significant negative correlation between DTF and TRFO was only found in STH15, and no significant correlation was found between these two traits in CSSF14, nor in CSSF15. Vegetative period (VP) was positively correlated with TSW in CSSF14, but the correlation between VP and the two soluble carbohydrates was not significant in CSSF14 (Table 6.1). In CSSF15, none of the three seed quality traits showed significant correlations with VP; but in STH15, all the seed quality traits were highly significantly negatively correlated with VP (Table 6.1). RP showed no significant correlation to any seed quality trait among environments, with only the one exception of a significant negative correlation to TSW in CSSF14 (Table 6.1). Plant height (PH) and

TSW were positively correlated in CSSF13, CSSF14 and CSSF15 while the correlation was not significant between these two traits in STH15 (Table 6.1). There was no significant correlation found between PH and sucrose or TRFO concentration in any environment (Table 6.1).

6.3.2 Mapping the monogenic traits on the linkage map of LR-26

Although four qualitative traits segregated in a monogenic manner among the LR-26 RILs, it was only possible to map cotyledon colour - on LG1 at 104.06 cM. It was not possible to map the other three traits unless they were placed at the end of the linkage groups with a significant gap.

6.3.3 QTL detection for each trait by environment

For DTE, significant QTLs were detected for all the three environments (CSSF13, CSSF14 and STH15; Figure 6.1). Consistent QTL regions located on linkage group (LG) 1 at around 200 cM could be found in all three environments (206.49-207.57 cM at CSSF13 and CSSF14, and 201.04-203.74 cM at STH15; Figure 6.1), explaining approximately 12.2-14.9% of the variability in this trait and with the early locus coming from the Lcu parent (Appendix E). A second significant QTL for DTE across environments was found on LG1 at approximately 80 cM. This QTL was located at 80.89-81.68 cM in CSSF13 and STH15; Figure 6.1), and this QTL explained approximately 10-18 % of the variability. There was a lower confidence QTL about 15 cM away on LG1 at 96.79-98.8 cM in CSSF14 (Figure 6.1) which explained about 5% of DTE-CSSF14 variance. Two unique DTE QTLs were detected from a single-environment: on LG3 at around 65 cM in CSSF13 (explaining approximately 5 % of DTE-CSSF13 variability), and on LG6 at around 55 cM in CSSF 14 (explaining approximately 15 % of DTE-CSSF14 variability) (Figure 6.1).

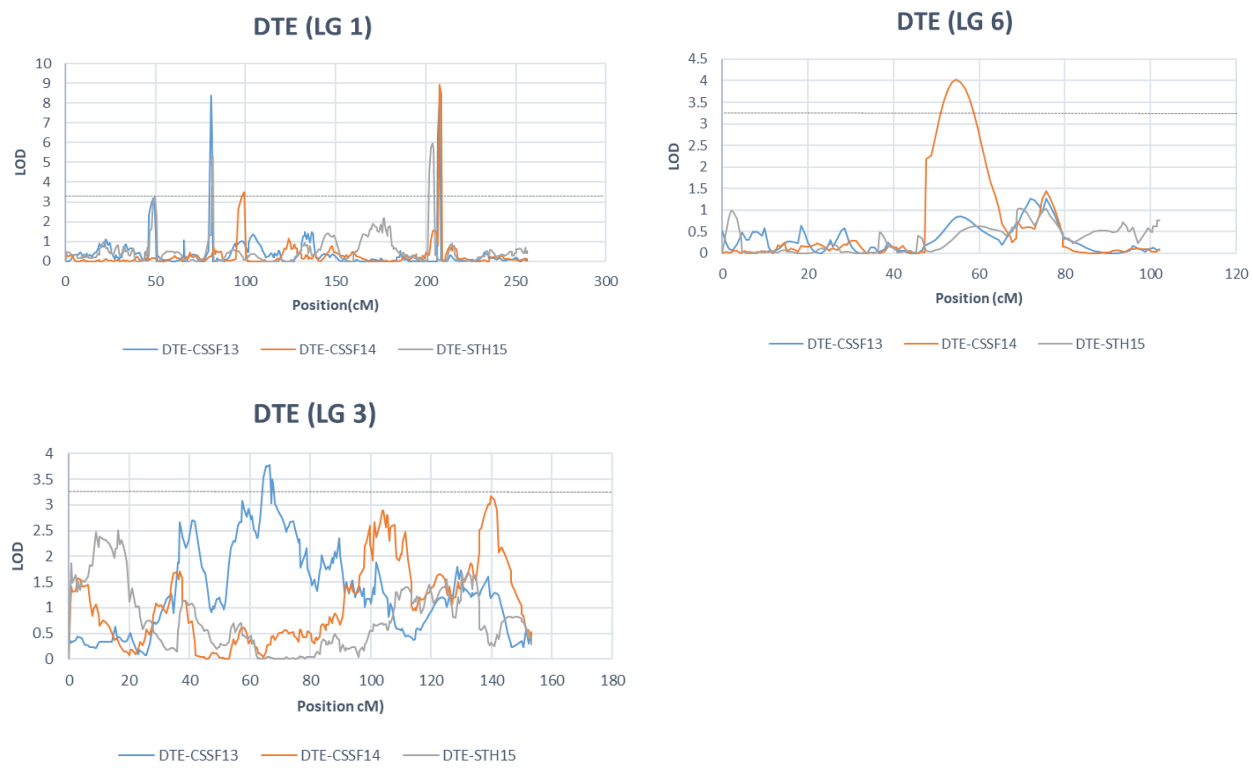


Figure 6.1 Days to emerge (DTE) QTL results for population LR-26 (Eston x IG 72815) phenotyped in three environments: Crop Science Seed Farm (CSSF) in 2013 (DTE-CSSF13), CSSF in 2014 (DTE-CSSF14) and Sutherland farm (STH) in 2015 (DTE-STH15). Dotted lines indicated significance LOD threshold at 3.3. LG: Linkage group; The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

For DTF, significant QTLs were detected for three environments: CSSF13, CSSF14 and STH15 (Figure 6.2) on LGs 1, 3, 4 and 5. No significant QTL for DTF were detected for CSSF15. A major effect QTL was found for all four environments on LG1 around 200 cM (201.04-203.74 cM; Figure 6.2), at the same region as one for DTE, and explained an average of 11.52% of the variance in DTF. A significant QTL was found in CSSF13 and CSSF14 on LG4 at 7.28-10.47 cM (Figure 6.2). This QTL explained approximately 7% of DTF variance at CSSF13 and 10% of DTF variance at CSSF14. Another significant QTL for DTF was found in CSSF14 and STH15 on LG 5 at 139.26-151.71 cM (Figure 6.2), explaining on average 6.2% of the variance in DTF. In all of these QTL, the early allele came from the Lcu parent.

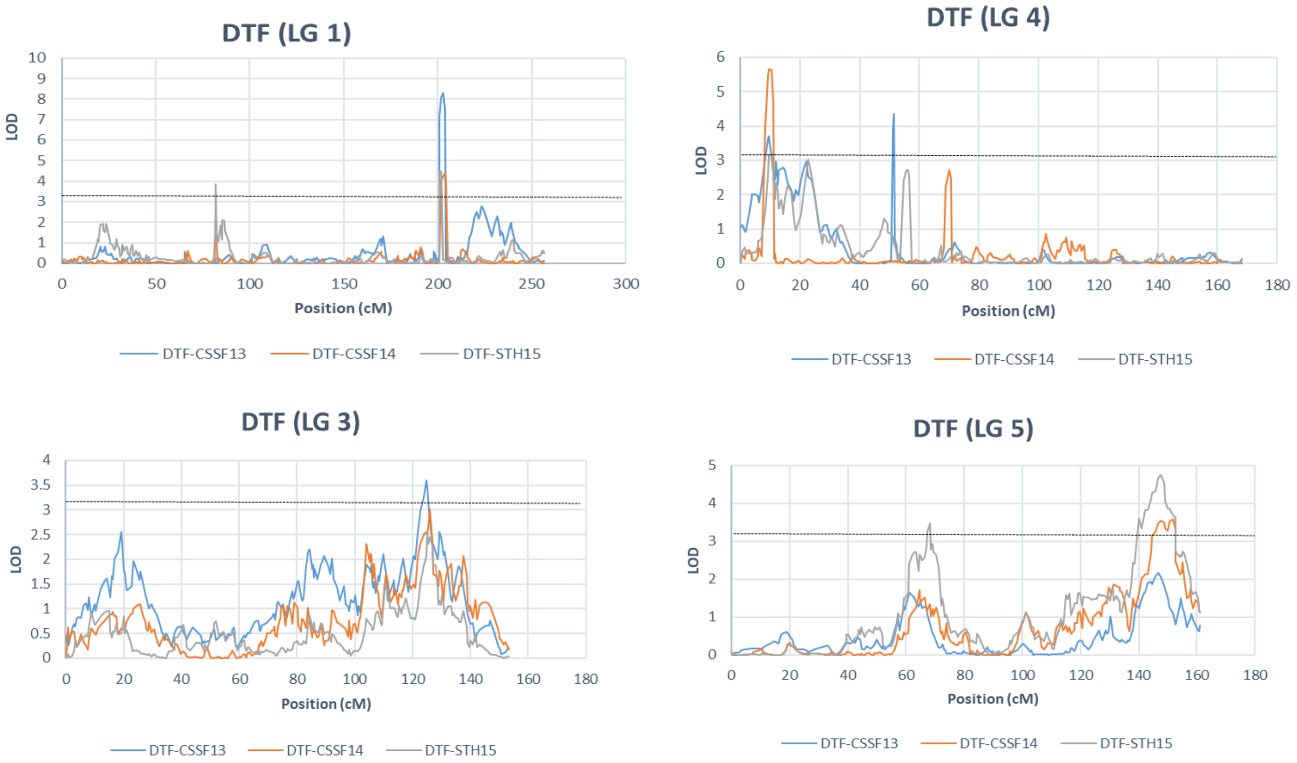


Figure 6.2 Days to flower (DTF) QTL results for population LR-26 (Eston x IG 72815) phenotyped in three environments: Crop Science Seed Farm (CSSF) in 2013 (DTF-CSSF13), CSSF in 2014 (DTF-CSSF14) and Sutherland farm (STH) in 2015 (DTF-STH15). Dotted lines indicated LOD at 3.2. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

Significant QTLs for VP were detected on LGs 1, 4, 5 and 7 (Figure 6.3), however, they were not consistent across environments; with the exception of the QTL on LG7 at approximately 120 cM (Figure 6.3) which was shared in both CSSF15 (120.47-122.73 cM) and STH15 (124.27-124.57 cM), and explained 7.7% - 12.4% of the variance in this trait. The highest effect QTL for VP was found in STH15 on LG5 at 67.18-67.48 cM, and explained more than 16% of the variance at this location. The donor of shorter VP in CSSF15 came from Ler; and shorter VP of other environments (CSSF13, CSSF14 and STH15) came from Lcu.

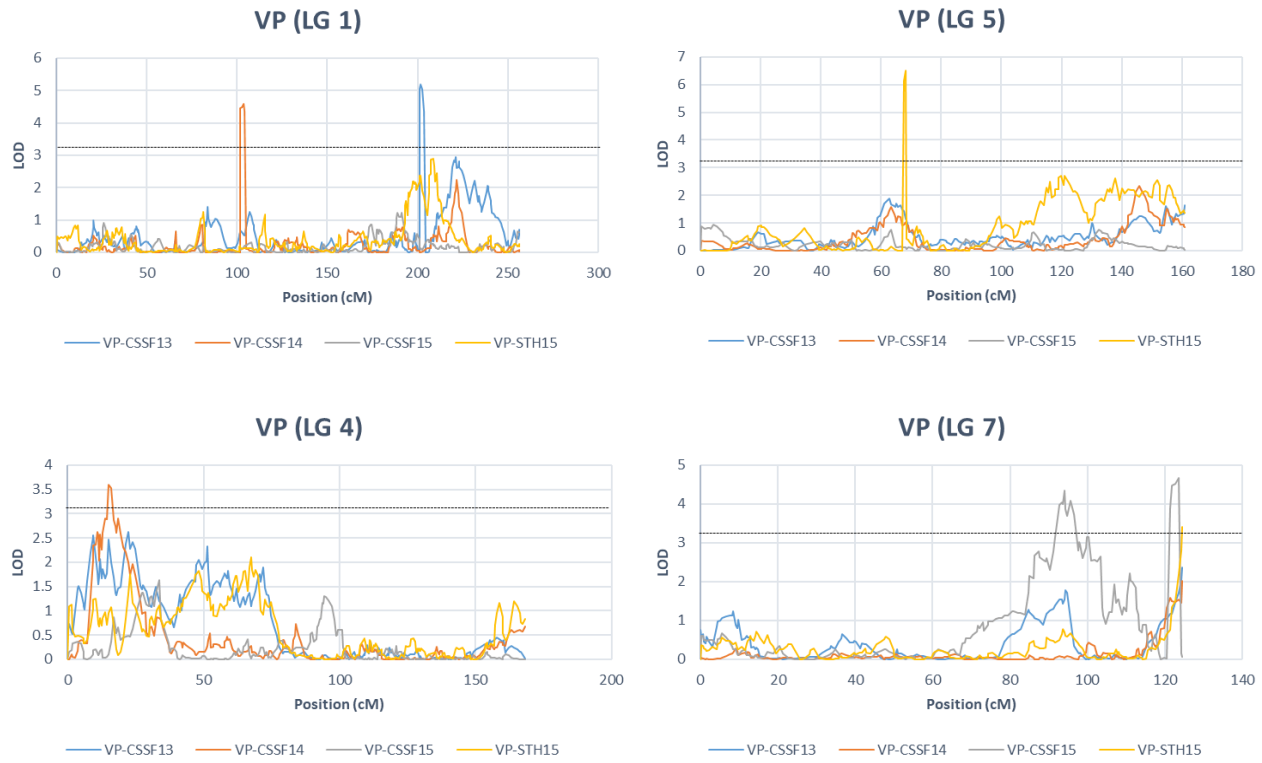


Figure 6.3 Vegetative period (VP) QTL results for population LR-26 (Eston x IG 72815) phenotyped in four environments: Crop Science Seed Farm (CSSF) in 2013 (VP-CSSF13), CSSF in 2014 (VP-CSSF14), CSSF in 2015 (VP-CSSF2015) and Sutherland farm (STH) in 2015 (VP-STH15). Dotted lines indicated LOD at 3.2. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

For RP, there was only one significant QTL detected, and only in STH15 (Figure 6.4). This was detected on LG5 at around 154-155 cM and explained 10.4% of the variance. The donor of shorter RP in STH15 came from Lcu.

Significant QTLs for PH were detected across all four environments on LGs 1, 3, 4 and 7 (Figure 6.5). A major effect QTL of PH on LG1 at 83.18-85.32 cM was found across all environment except STH15 (Figure 6.5). This QTL is the only significant QTL for CSSF15 and explains on average 14% of the variance in PH in these environments. Another multi-environment PH QTL was found on LG1 at approximately 160 cM in CSSF13 and CSSF14 (Figure 6.5). The only significant QTL for STH15 near there was at 133 cM on LG1 (Figure 6.5). In all of these QTL, the taller PH allele came from the Lcu parent.

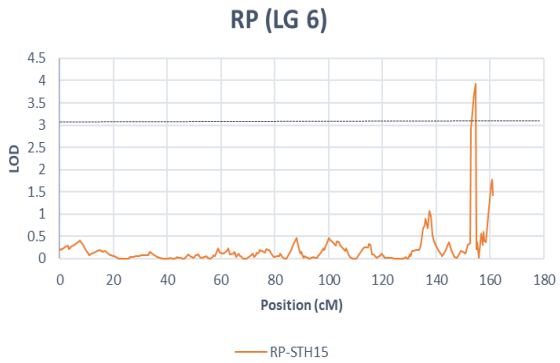


Figure 6.4 Reproductive period (RP) QTL results for population LR-26 (Eston x IG 72815) phenotyped in Sutherland farm (STH) in 2015 (RP-STH15). Dotted lines indicated LOD at 3.1. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

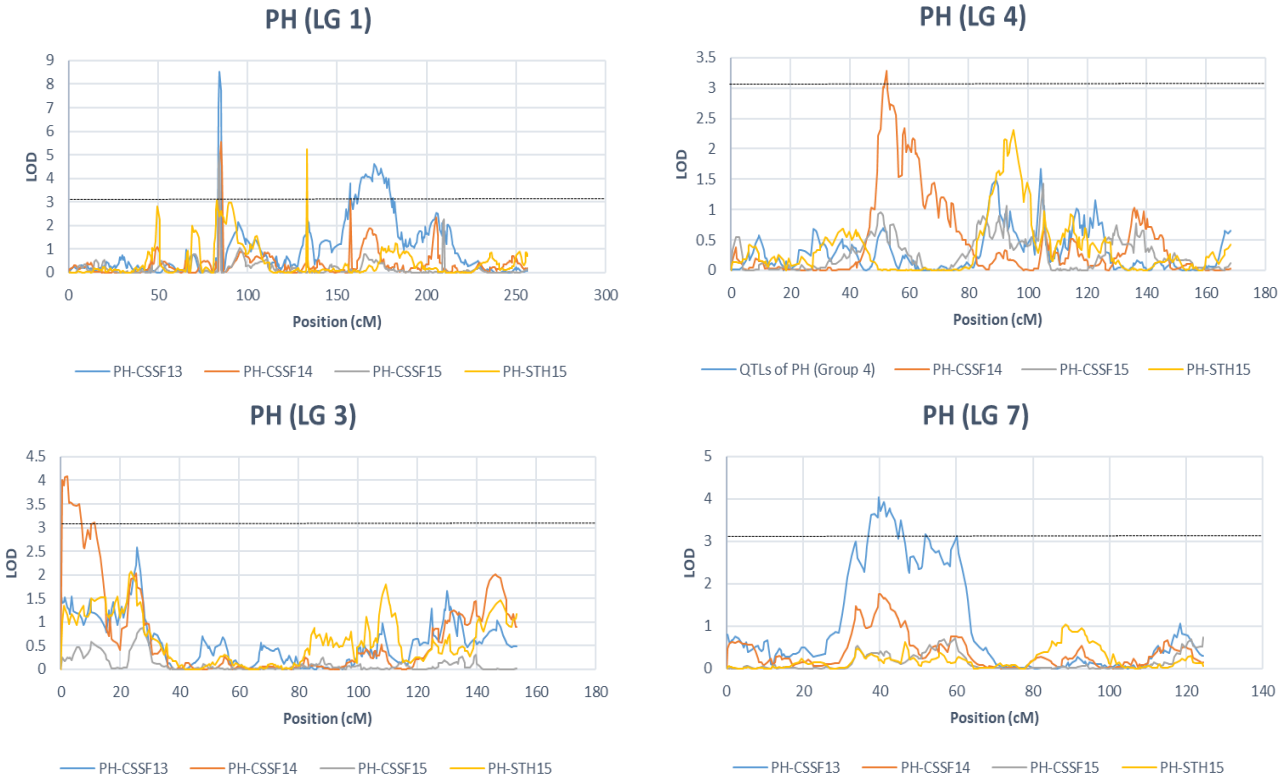


Figure 6.5 Plant height (PH) QTL results for population LR-26 (Eston x IG 72815) phenotyped in four environments: Crop Science Seed Farm (CSSF) in 2013 (PH-CSSF13), CSSF in 2014 (PH-CSSF14), CSSF in 2015 (PH-CSSF2015) and Sutherland farm (STH) in 2015 (PH-STH15). Dotted lines indicated LOD at 3.1. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

Significant QTLs contributing to variance in TSW were detected from all four tested environments and can be found on most of the LGs except for LG2 (Figure 6.6). A major effect QTL on LG1 at 201.04-208.92 cM was detected at all the tested environments and explained an average of 8% of the variance. Another consistent QTL for TSW across all tested environments was found on LG7 at around 70 cM (Figure 6.6) and explained an average of 4% of the variance among LR-26 RILs.

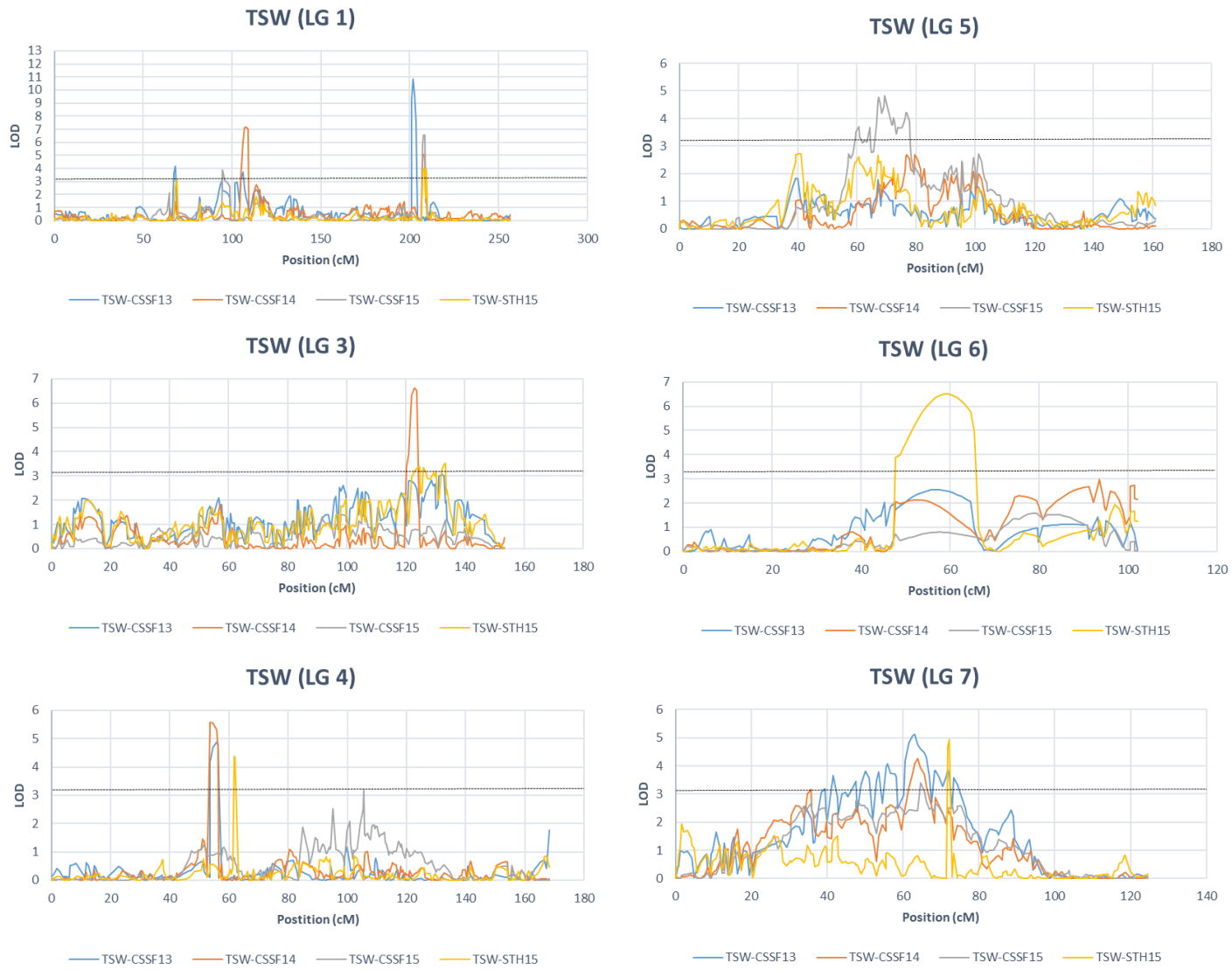


Figure 6.6 Thousand seed weight (TSW) QTL results for population LR-26 (Eston x IG 72815) phenotyped in four environments: Crop Science Seed Farm (CSSF) in 2013 (TSW-CSSF13), CSSF in 2014 (TSW-CSSF14), CSSF in 2015 (TSW-CSSF2015) and Sutherland farm (STH) in 2015 (TSW-STH15). Dotted lines indicated LOD at 3.2. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

Another consistent QTL for TSW was found on LG4 from 52.64-61.84 cM (Figure 6.6) in most environments except for CSSF15 and explained an average of 4.76% of variance in the three environments. A QTL was found on LG1 at around 104.06-109.01 cM in both CSSF13 and CSSF14, and coincided with the cotyledon colour marker (Figure 6.6) and explained 6.7% of variance in CSSF13 and CSSF14. A significant QTL in CSSF15 was located around 10 cM away on LG1 at 94.1 cM (Figure 6.6). This QTL explained 6.5% of variance of at this location. The larger TSW allele came from the Lcu parent.

Significant QTLs for the concentration of sucrose in seeds from three environments were detected on LGs 1, 3 and 4 (Figure 6.7). A major QTL in CSSF14 was found on LG1 at 199.72-200.04 cM and it explained 15% of sucrose variance at this environment. The significant QTL on LG3 at around 65 cM was detected in both CSSF14 (63.09-64.4 cM) and STH15 (66.52 cM), and this QTL explained an average of more than 8% of the sucrose variance at these two environments. A major QTL was found in STH15 on LG4 at around 51 cM which explained nearly 20% of the variance in sucrose concentration in this environment. The higher sucrose concentration allele came from the Lcu parent. For seed TRFO concentration, significant QTLs were only found in two environments in 2015 (CSSF15 and STH15) (Figure 6.8). For CSSF15, significant QTLs were found on LGs 1, 4, 5, 6 and 7. In STH15, significant QTLs were detected on LGs 3 and 7. The major effect QTL on LG 7 at approximately 30 cM was shared by both environments (Figure 6.8), and this QTL explained an average of 10.5% of variation in TRFO across both environments. The higher TRFO concentration allele came from the Lcu parent.

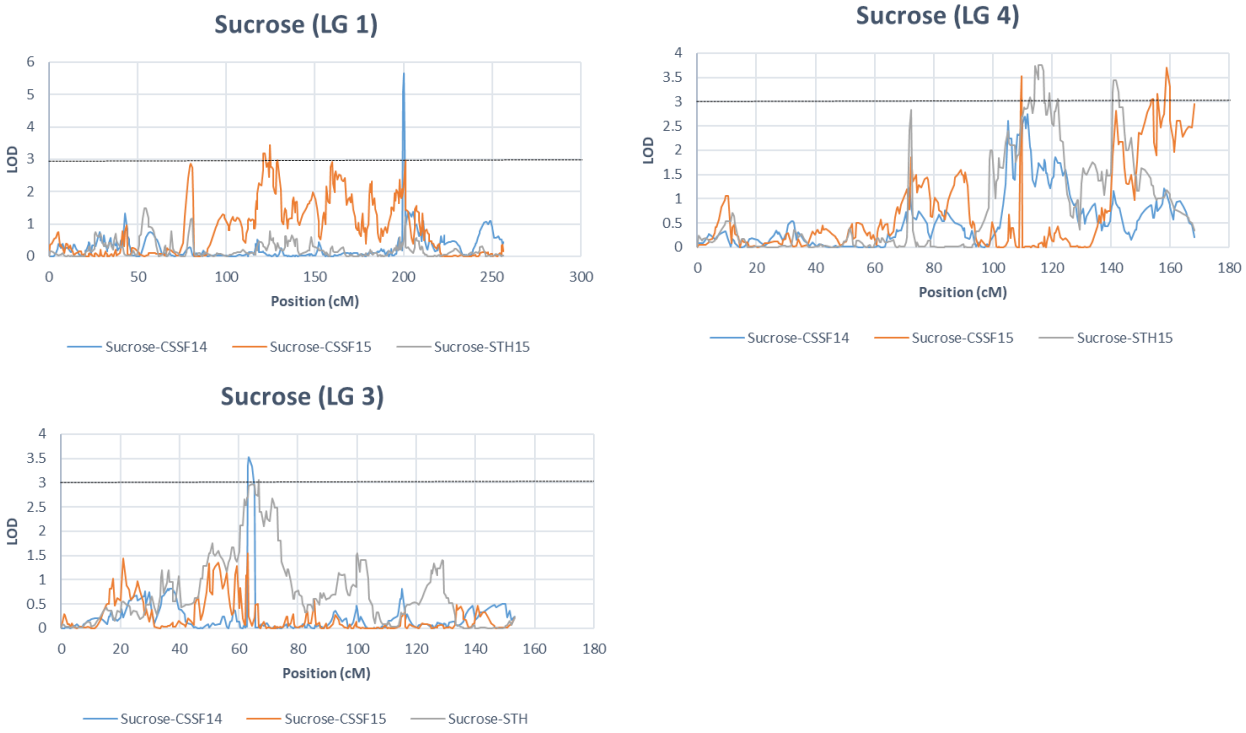


Figure 6.7 Sucrose concentration (Sucrose) QTL results for population LR-26 (Eston x IG 72815) phenotyped in three environments: Crop Science Seed Farm (CSSF) in 2014 (Sucrose-CSSF13), CSSF in 2014 (Sucrose-CSSF15) and Sutherland farm (STH) in 2015 (Sucrose-STH15). Dotted lines indicated the LOD threshold at 3. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

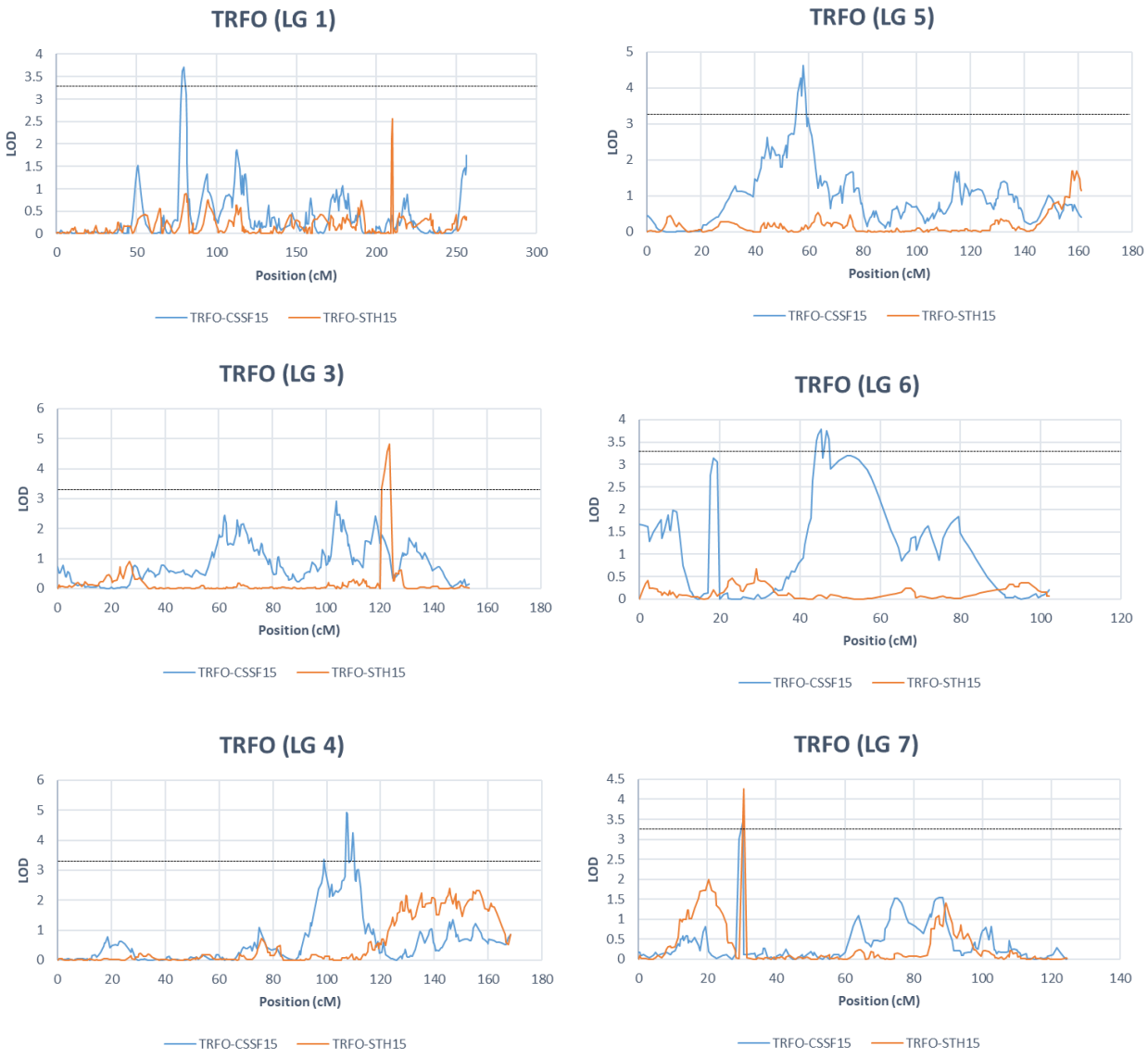


Figure 6.8 Total raffinose family oligosaccharides concentration (TRFO) QTL results for population LR-26 (Eston x IG 72815) phenotyped in three environments: Crop Science Seed Farm (CSSF) in 2015 (TRFO - CSSF15), and Sutherland farm (STH) in 2015 (TRFO -STH15). Dotted lines indicated the LOD threshold at 3.3. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

6.4 Discussion

In this study, the phenotypic variation within all the traits recorded in the four different environments (Chapters 3 and 4) were further combined with the LR-26 genotypic results using the linkage map described in Chapter 5. The objective of the work described in this chapter was to perform a genotype-phenotype association analysis using these data.

In this chapter, a correlation coefficient test was first used to help clarify the relationships among the traits. This was done in individual environments because of the significant GxE interactions reported in the previous chapters. Based on Spearman's correlation test results (Table 6.1), significant correlation was found between most pairs of phenology traits. For instance, a positive correlation was found between DTE and DTF while DTE was negatively correlated with VP (Table 6.1). Physiologically, the flowering time of lentil is controlled by the photo-thermal threshold during the vegetative period (Summerfield et al., 1985; Roberts et al., 1986). Since the timing of flowering of lentil is related to a specific critical photo-thermal condition, the flowering of genotypes was most likely to occur at a similar time of the year rather than a specific number of days, and, therefore, a negative correlation between DTE and VP results (Table 6.1). The results here reflected that the days to germinate and the needed vegetative period varied among the genotypes. And DTF was affected by emergence time and therefore a positive correlation between DTE and DTF.

The relationship of RP with the earlier phenology traits was the least consistent of the correlations (Table 6.1). The drying of legume pods, an indicator of maturity and therefore a determinant of RP, is affected by the climatic conditions (Davies, 1945). The RP in CSSF14 was more stretched out than in both CFFL15 and STH15 (Chapter 3, Figure 3.3) likely because of the cooler and more humid period during the maturation stage (Appendix A).

A relationship between phenology traits and PH was also detected in this population. A longer vegetative period may extend the period of shoot growth and this was seen in CSSF14 and CSSF15. Previously, a highly significant positive correlation between PH and DTF was reported in cultivated lentil (Hamdi et al., 1991). At STH15, however, there was a negative correlation between DTF and PH (Table 6.1). This was likely because the delayed emergence time resulted in a shorter VP so less time to add to height before the reproductive period started.

Consistent positive correlations between the three seed quality traits measured were also found within the multiple environments (Table 6.1). For instance, there was a consistent, highly significant, positive correlation ($p<0.001$) between TSW and both sucrose and TRFO (Table 6.1). This positive correlation may be attributed to the fact that the latter two are important soluble sugars and major storage carbohydrates in plant seeds (Sengupta et al., 2015), and seed size and seed weight are highly influenced by the mass of the various storage compounds in plant seeds (reviewed in Dante at al., 2014). Since RFOs are a type of galactosyl-sucrose oligosaccharides, and sucrose is the precursor of RFO biosynthesis (Peterbauer and Richter, 2001), these two soluble sugars showed strong positive correlation within all environments (Table 6.1). A positive correlation between seed TRFOs level and sucrose concentration was also reported by Tahir et al. (2011) in cultivated lentil.

Generally, it is agreed that as major storage carbohydrates in legume seeds, RFOs and sucrose are possible sources of energy for seed germination (Kuo et al., 1988); however, the physiological role of RFOs on plant growth regulation is still debatable (reviewed in Sengupta et al., 2015). While RFO has been reported to be essential for early germination in Arabidopsis (Gangl and Tenhaken, 2016) and pea (Blöchl et al., 2007), in soybeans (Dierkinga and Bilyeu, 2009) and chickpeas (Gangola et al., 2016), RFO did not appear to be necessary for germination, as there may be other sources of energy of the seeds. In LR-26, a consistently negative correlation existed between DTE

and TRFO, DTE and sucrose concentration, as well as DTE and TSW (Table 6.1). Thus, while the role of RFOs in lentil remains unclear, these consistently significant correlations suggest that the level of these seed storage carbohydrates as well as the total seed mass may have an impact on the early germination in these RILs and perhaps a breeding objective of lowering TRFO should be undertaken with caution.

There were four morphology and seed quality traits segregating in a monogenic manner in LR-26, but only cotyledon colour could be mapped - on LG1. Fedoruk et al. (2013) also mapped cotyledon colour in a Lcu RIL population on their LG1. In both Fedoruk et al. (2013) and this study, this corresponds to chromosome 1 of the lentil reference genome.

It was not possible to map three other qualitative segregating traits - flower colour, pod dehiscence and hilum colour - in the LR-26 linkage map. Since it is a really high-density map of LR-26 (average marker interval < 1 cM), the chance that the markers associated to the qualitative variations were not segregating in the RILs was low. However, while there were only two phenotypes being scored in the population for these qualitative traits, the level of flavonoids accumulation controlling the expression of flower colour or hilum may be controlled by multiple loci. Additionally, the complete pod shattering mechanism has also been reported to be under a complex scenario involving multiple genes in soybeans (Dong et al., 2014) and other legumes (Ballester and Ferrándiz, 2017). Though the rating of pod dehiscence was based on the state of lignification, the possible existence of other mechanism may still bias the scoring of this trait. Therefore, the strategy for genetic mapping of these traits needs to be re-evaluated in this population.

To further help unravel the genomic position and possible genetic linkages among the traits, the phenotypic means of the quantitative traits and the genetic map of LR-26 were used for marker-trait association analysis for each environment. The QTL mapping results suggested that the quantitative agronomic and seed quality traits are

each affected by multiple loci (Figures 6.1 - 6.8). The QTL mapping results help explain the phenotypic performance of these traits. The ANOVA result for genotypic effect on RP was not significant at CSSF14. Although the genetic effect was found to be slightly significant at CSSF15 ($p<0.05$) (Table 3.6), the characterized maturation time might still be masked by multiple environmental effects. However, as was discussed Chapter 3, the rating of RP was rather difficult which decrease the quality of the phenotypic data for detecting significant QTL in CSSF15. And only a single QTL from only one location - STH15 - was detected for RP. This fits with the strong environmental effect on the phenotypic variation of RP noted in LR-26 (Chapter 3).

QTL co-localization of various traits can help to identify possible genetic linkages or pleiotropy, and the results may help explain the correlation between traits in LR-26 (Table 6.1). For example, by comparing the QTLs of all the agronomic traits (Figure 6.9), DTE and DTF both have their major effect QTL located at approximately 200 cM on LG1. This locus reflects the highly significant correlations between DTE and DTF in CSSF13, CSSF14 and STH15 (Table 6.1). This QTL on LG1 was also significant for another important phenology trait for measuring flowering time of VP at CSSF13 (Figure 6.9). The calculation of VP was a result of the subtraction between DTE and DTF. While DTE and DTF shared several QTLs consistently, there were no other QTL at other environments shared between VP and DTE or VP and DTF. Again, this indicated that VP was truly representing the flowering time in the RILs, while DTF was affected by the high variability of emergence time in this population.

Furthermore, by superimposing the QTLs of all tested agronomic traits (Figure 6.9), another QTL hotspot can be found also on LG1, at around 85 cM. This QTL region was found to be significant for DTE at CSSF13 and STH15, DTF at STH15, and PH at CSSF13, CSSF14, and CSSF15 (Figure 6.9).

The significant QTLs of the tested seed quality traits could be detected from most of the linkage groups except LG2 (Figure 6.10), likely because LG2 was a short linkage

group - only 22.73 cM in length (Table 5.1, Chapter 5). The genetic control of these seed quality traits appeared to be mostly independent. A seed quality QTL hotspot region was found on LG1 at around 100 cM. This spot was significant to TSW at CSSF14 and CSSF15. This TSW QTL was between the other two seed storage carbohydrate QTL on LG1, of which the one for TRFO at CSSF15 20 cM away at around 80 cM on LG1, and a QTL for sucrose, also at CSSF15, was located 25 cM away at around 125 cM on LG1. And there was only one significant QTL shared by sucrose (at CSSF15 and STH15, this QTL only explained approximately 9 % of the sucrose variance at each environment) and TRFO (at CSSF15, this QTL explained approximately 11 % of the TRFO variance) which was located on LG4 at around 110 cM (Figure 6.10).

It is important to note, however, that these QTL hotspots both correspond to very long stretches of the lentil reference genome involving many genes. The QTL results were also derived from only one bi-parental population, with limited population size. Additionally, as was discussed in Chapter 5, possible pseudolinkage between markers from chromosomes undergoing structural rearrangement may disturb the QTL results. For example, the QTL region on LG1 at approximately 200-210 cM was also found significant for TSW at all environments (Figure 6.6), hence was shared by several phenology traits and seed quality traits. Considering the odds these traits sharing same genetic controlling is low, the co-localization of certain QTLs is more likely based on several independent QTLs being mapped to the same location. The LG1 of this map was generated by markers associated with three separate Lcu chromosomes. Therefore, this hotspot of co-localized QTLs among various traits may refer to several independent QTLs that were linked in this population, hence, further genetic analysis would be needed to clarify the genetic regions underlying these QTLs.

Cotyledon colour was mapped at 104.06 cM on LG1. The locus of cotyledon was linked to a significant QTL for TSW-CSSF13 and TSW-CSSF14, and about 10 cM away from a significant QTL for TSW-CSSF15 (Figure 6.6). This result indicated a possible

linkage between cotyledon colour and seed weight in the individuals of LR-26. Abbo et al. (1991) reported a significant association between cotyledon colour and seed weight in F_2 populations that were derived from Lcu crossed to its wild progenitor species, *L. orientalis*. And Fedoruk et al. (2013) reported genetic linkage between cotyledon colour and two seed size indicators - seed diameter and seed plumpness in a lentil RIL population. On the other hand, Khazaei et al. (2017) found that there was no association between cotyledon colour and several seed size indicators, such as seed diameter, seed plumpness and seed thickness, based on the study involving 138 accessions of cultivated lentils suggesting the two traits are linked but not tightly. Since the mapping population was derived from limited genetic sources of either red and small seed (Ler) or yellow and medium to large seed (Lcu), these two traits (TSW and cotyledon colour) may not segregate independently in the 167 RILs of LR-26. Fedoruk et al. (2013) also reported a linkage between cotyledon colour and DTF. In LR-26, cotyledon colour 104.06 (at LG1, 104.06 cM) was also very close to a QTL of VP-CSSF14 (at LG1, 101.52-103.83 cM), which explained 12.4% of the VP-CSSF14 variance.

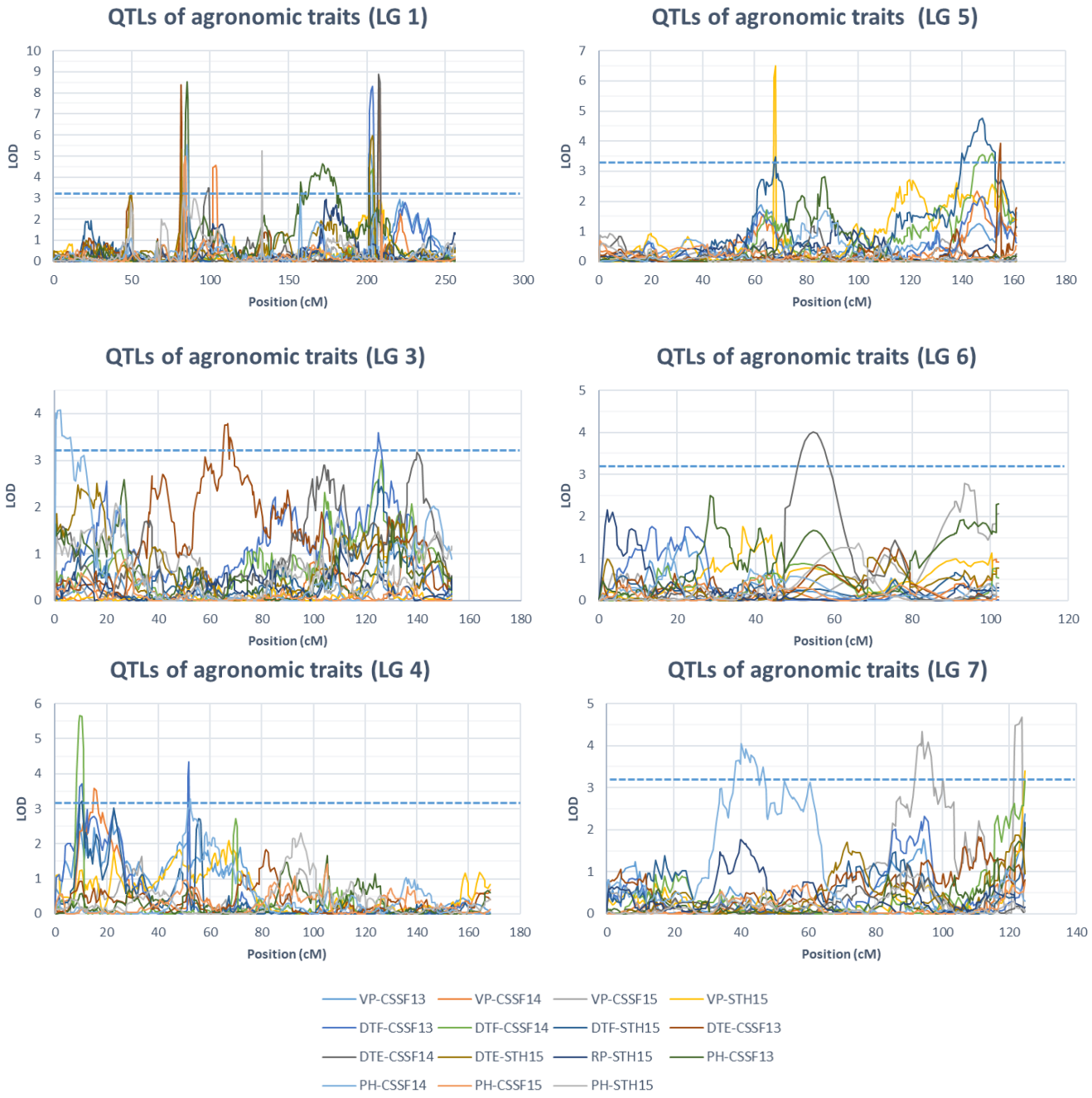


Figure 6.9 Genome-wide QTL detection results for population LR-26 (Eston x IG 72815) of tested agronomic traits phenotyped at four environments. The traits were days to emerge (DTE); vegetative period (VP); days to flower (DTF); reproductive period (RP); and plant height (PH). The tested environments were Crop Science Seed Farm (CSSF) in 2013 (CSSF13), 2014 (CSSF14), and 2015 (CSSF15); and Sutherland farm (STH) in 2015 (STH15). Only the environments with significant QTL detected on tested traits are displayed. The results were displayed for separated linkage groups based on the LR-26 linkage map. There were no significant QTL detected on linkage group 2 for any of the traits. LOD: logarithm of odd ratio. Dotted blue lines highlighted LOD value at 3.2. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

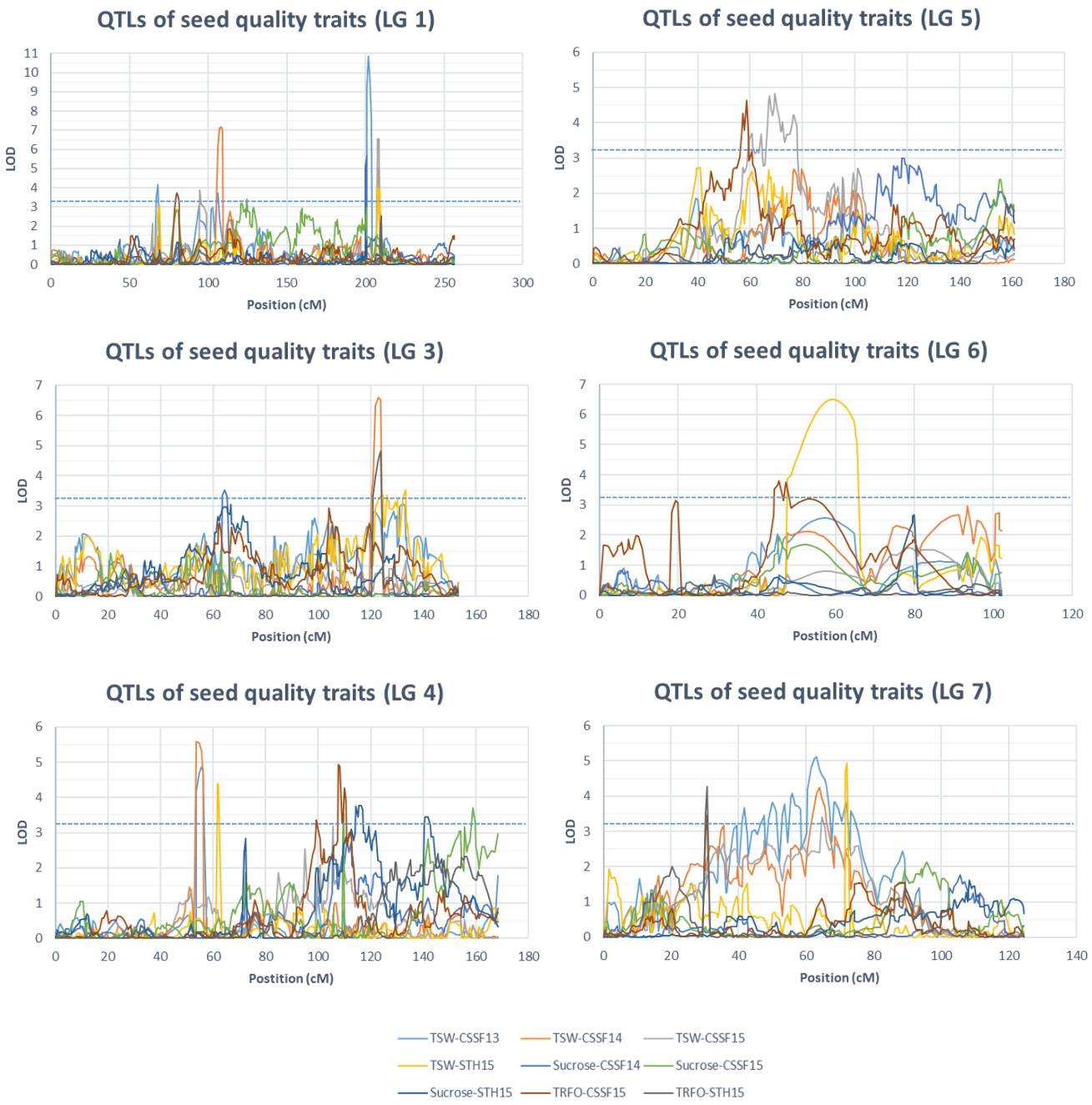


Figure 6.10 Genome-wide QTL detection results for population LR-26 (Eston x IG 72815) of tested seed quality traits calculated at four environments. The traits were thousand seed weight (TSW); sucrose concentration (Sucrose); and total raffinonse family oligosaccharides concentration (TRFO). The tested environments were Crop Science Seed Farm (CSSF) in 2013 (CSSF13), 2014 (CSSF14), and 2015 (CSSF15); and Sutherland farm (STH) in 2015 (STH15). Only the environments with significant QTL detected on tested traits are displayed. The results were displayed for separated linkage groups based on the LR-26 linkage map. There were no significant QTL detected on linkage group 2 for any of the traits. LOD: logarithm of odd ratio. Dotted blue lines highlighted LOD value at 3.2. LG: Linkage group. The QTL detection was done based on a LR-26 linkage map developed using SNP markers detected using genotyping-by-sequencing.

This is the first QTL mapping study of a Lcu x Ler introgression population, and the high-density linkage map provided the opportunity to locate genomic regions and allele donors of many phenotypic traits of interest for future studies.

To develop an efficient introgression-by-design pipeline for pre-breeding, breeders seek genomic information to help manipulate the direction of selections. Through the association of SNP variants and phenotypic variations from multi-environmental field trials using an Lcu x Ler derived interspecific RIL population, researchers can identify the genetic regions underlying the phenotypic variability from the introgression. QTL results based on polymorphic SNP variants help develop an understanding the genotype-phenotype associations within the population. Using these genome-wide high-density polymorphic SNP variants, QTL regions were significantly associated with most of the tested phenotypic variations and were detected across the whole genome. The overlapping of significant QTLs among traits also help illustrate the underlying causes of phenotypic correlations in these RILs. The QTL hotspots identified from only one bi-parental population may not fully reflect the linkage between QTLs, and the limited population size also decreased the power of QTL detection. To utilize the genotype-phenotype association analysis with additional genetic populations in the future may help improve the result of this study.

CHAPTER 7. GENERAL DISCUSSION

7.1 Research summary

The purpose of this study was to investigate the impacts and challenges of CWI in the Lcu x Ler derived RILs at both phenotypic and genomic levels. The combination of information provides the opportunity to look at the genomes of Lcu x Ler derived progenies. Previously, most studies mapping candidate QTLs using interspecific genetic populations focused on traits related to biotic stress resistance, abiotic stress tolerance or enhancing yield; with rather fewer studies focused on the agronomic trait or nutritional or compositional quality implications of the introgressions (Dempewolf et al., 2017; Fernie et al., 2006; Hajjar and Hodgkin, 2007). This study, however, was focused on exploring variations in a wide range of traits from phenology and morphology to seed quality and seed compositional traits caused by introgression of wild genetic material. To investigate the genetic control underneath the phenotypes, the most high-density genetic map of an interspecific lentil population so far, was developed.

The first hypothesis in this study was that significant genetic effects that were caused by CWI could be found to affect the phenotypic variation of specific agronomic and seed quality traits. According to the test results from multi-environment field trials reported in Chapters 3 and 4, the interspecific RILs reacted differently to specific environments as significant GxE interaction was detected in each tested trait, and significant genetic effects on the phenotypic variations of all tested traits were shown.

The second hypothesis was to test if the phenotype of each trait segregated predictably in the tested environments, and that the genome would be evenly distributed between Lcu and Ler genomic fragments. According to the phenotypic frequency distribution results of both LR-26 and LR-59 in Chapters 3 and 4, as well as the allele distribution frequency from the GBS result of LR-26 in Chapter 5, the

phenotypic distributions of most traits were skewed towards the Lcu parent in both populations, while the allele distribution in most of the chromosomes also skewed to the cultivated parent in LR-26. This hypothesis was therefore rejected.

The third hypothesis was that marker-trait associations could be used to identify Ler introgression regions of importance to the tested agronomic and seed quality traits. To do this, QTL analysis was carried out using a multiple QTL mapping model (Jansen and Stam, 1994) in LR-26. Because of the significant GxE interaction detected among all the tested traits, the analysis was done by the trait value of each single environment. The results helped identify the allele donor of each QTL underlying the variation of the examined traits, including the agronomic and seed quality traits. Exceptions were for traits with low heritability where the environmental impact was too large, such as RP of CSSF14 and CSSF15, and therefore significant QTLs were not detected in such cases. The QTLs were found across the whole genome and helped demonstrate the phenotypic correlation between traits being observed.

7.1.1 Phenotypic analysis of the segregating traits among two Lcu x Ler derived RIL populations, LR-26 and LR-59

In this study, the goal was to investigate impacts of CWI on various complex traits, beyond disease resistance, in LR-26 and LR-59. Certain traits involved in plant phenology are very important indicators of adaption of lentil (Erskine et al., 1990) and were characterized in LR-59 and LR-26; these traits included seed emergence time (in DTE), flowering time from seeding (in DTF) and flowering time from plant emergence (in VP), and time from plant flowering to maturation (in RP). Plant morphological variables were also characterized in the RILs, these morphological traits included PH at flowering stage and flower colour.

While most CWI studies have focused on specific beneficial traits such as disease resistance or abiotic stress tolerance, or even yield improvement, the effect of utilization

of CWRs on seed or fruit quality traits remained less exploited and therefore has only started attracting more attention recently (Dempewolf et al., 2017; Fernie et al., 2006; Eshed and Zamir, 1994; Hajjar and Hodgkin, 2007; Kaushik et al., 2017). In my study, impacts of introgression were found on seed quality traits in LR-59 and LR-26. In both interspecific RIL populations, two major lentil classification indicators of seed size (in TSW) and seed cotyledon colour, as well as two seed soluble carbohydrates concentration (sucrose and TRFO) were also found segregating among the RILs.

Overall, the *in situ* phenotypic characterization of the interspecific progeny is arguably the most important step in pre-breeding to unravel the potential of wild introgression (Kaushik et al., 2016; Sharma et al., 2013). This study established field trials of the interspecific populations in the Western Canadian lentil growing environment for multiple site-years. Through field observations, the goal of identifying the phenotypic variability and stability, as well as the potential challenges was met.

7.1.2 Genotypic linkage analysis of the mapping population of LR-26

This study provided insight into possible challenges of introgression mapping in lentil interspecific populations. Several issues associated with the wide-cross between Lcu and Ler have been previously reported and further discussed in this study. For instance, divergence between the two parental species is associated with a major chromosomal rearrangement due to a reciprocal translocation (Bhadauria et al., 2017; Gujaria-Verma et al., 2014). This results in aberrant non-homologous pairing within Lcu x Ler derived hybrids (Fiala et al., 2009; Ladizinsky et al., 1985). Additionally, segregation distortion has been reported for genomic segments in the progeny of Lcu x Ler crosses (Zamir and Tadmor, 1986) and this could lead to distortion of phenotypic values from expectation. The impacts of these issues on plant performance were observed in the field trials of LR-26 and LR-59 as low fertility which may be attributed to semi-sterility, translocation heterozygotes, and phenotypic population distortion.

To illustrate the impact of Lcu x Ler hybridization at the genome level, further assessment of the genotype was done through high-throughput genotyping using the larger and more carefully developed, RIL population of LR-26. The linkage analysis results helped identify the problematic genomic regions associated with the previously observed issues. A larger linkage group was formed due to the pseudolinkage of the SNPs from multiple chromosome including parts of chromosomes 1 and 5 that were involved in the reported reciprocal translocation in Lcu (Bhadauria et al., 2017; Gujaria-Verma et al., 2014), as well as a putative large tandem duplication located in chromosome 7 of Lcu genome (Larissa Ramsay, personal communication). This pseudolinkage made the further identification of candidate genes more difficult. Furthermore, the limitation on maintaining the size and variability of an interspecific Lcu x Ler population during development also decreased the resolution of the QTL linkage analysis study.

The genotype-phenotype associations within LR-26 helped identify significant QTLs among agronomic and seed quality traits based on single given environment were detected across the whole genome of LR-26. The results also indicated that these quantitative traits can be affected by several minor-effect loci in the RILs.

7.1.3 Assessing the Lcu x Ler derived RILs for lentil breeding

Before QTL analysis, the phenotypic association between traits were first combined to provide some insights into the relationship in the agronomic and seed quality variations in LR-26. For example, while selecting for a lower TRFO concentration may be desirable for reducing bloating effect, other than the consequence of a decrease of beneficial prebiotic element in lentil, a possible association of later emergence time, smaller seed size, and lower sucrose concentration may come along at the same time. After combining information from both phenotypic and genotypic analysis from interspecific RILs, some possible benefits and challenges associated with

the Lcu x Ler introgression were acknowledged in this study. For instance, the structural rearrangement of the genome during species divergence between Lcu and its distant relative species Ler may increase the difficulty of pre-breeding. The challenges included the permanent translocation heterozygosity, decreased sterility, and lower frequency of crossover between the multivalent chromosomes at meiosis. These issues made the field-based phenotyping more difficult as the viability and uniformity of plants within a genotype were affected. They may also result in linkage drag within introgression segments from the rearrangement region. Therefore, further caution is recommended in order to incorporate these interspecific materials into the breeding program. For instance, while the introgression brought novel variations such as lower seed sucrose and TRFO concentration than the current cultivars (Tahir et al., 2011b), it also brought some of the wild performance back to the individuals. The undomesticated traits such as seed dormancy (seen as late emergence) and pod dehiscence increased the work of management. Furthermore, the non-uniform performance of the RILs also made the data characterization more difficult.

To fully benefit from introgression, it is critical to maintain population size to include all the possible recombinations and genetic variability. In the case of LR-26 and LR-59, a starting population size of more than 300 lines at the F_2 stage (Tullu et al., 2013), after six generations of single seed descent and few more generations of self-pollination, the number of RILs declined to as low as 67 RILs for the earlier developed population of LR-59. Nonetheless, even though the second interspecific population of LR-26 was more carefully maintained, the population size was still reduced by nearly half to the 172 RILs that were included in this study. Furthermore, each generation of self-pollination of the Lcu x Ler, translocation heterozygotes would produce semi-sterile meiotic products. The reduction of interspecific population size leads to a loss of genotypic combination and further decrease the power of QTL analysis. Additionally, chromosomal rearrangements such as reciprocal translocations, as well as tandem

duplications, result in marker presudolinkage among these problematic regions and increases the difficulty of QTL mapping using the interspecific RILs.

Previously, Bhadauria et al. (2017) mapped 14 QTLs underlying the resistance of several fungal disease, including anthracnose race 0, anthracnose race 1, and stemphylium blight existed in Ler based on an intraspecific Ler RIL population LR-66 (derived from a single F1 plant of accessions L01-827A and IG 72815 cross as described in Bhadauria et al., 2017). In Bhadauria et al. (2017), the QTLs with alleles contributed from L01-827A were detected outside of the chromosomes 1 or 5 but from chromosomes 2, 3 and 7, which may be easier for breeders to introgress the resistance without big segments of linkage drag. On the other hand, four disease resistance QTLs (two for anthracnose race 0 and two for anthracnose race 1) contributed by IG 72815 were located on chromosome 5. Through multiple generations of backcrossing, breeders may still use IG 72815 as resistance donor for the future, but it may require more careful management and would benefit from the use of marker-assisted selection.

7.2 Afterthoughts and suggestions for future research

7.2.1 Introgression breeding: bringing diversity back into modern crop species

The importance of the conservation of wild genetic resources for crop improvement has been emphasized since the pioneer work of the very influential Russian botanist Nikolai Vavilov (1887-1943). The use of CWR for genetic improvement has been tried in many crop species during the past century (Bessey, 1906) but was limited in use. Today, the potential of CWR in securing food production has been widely recognized, and successful examples can be found in many major crop species (reviewed in Dempewolf et al., 2017). For example, in tomato, there is a large genetic variation between elite cultivars and the local landraces (Corrado et al., 2013), as most elite tomato varieties today contain wild introgression segments contributing to disease resistance (Menda et al., 2014).

However, the requirement of a longer selection period and possible linkage drags of detrimental traits had deterred the usage of genebank resources for some traditional crop breeders. A strategy to select based on the genotype instead of the phenotype was proposed by Tanksley and McCouch (1997) to improve the efficiency of utilization of wild genetic materials. This approach aims to help guide breeders in the use of exotic materials. Since molecular markers can be used to track the introgression of desirable traits, the construction of molecular maps and the execution of marker-trait association analysis study using interspecific populations are crucial steps in completing the process.

As a reservoir of genetic diversity, CWR would be a sustainable source of novel variation. Furthermore, while CWR is mostly used as a direct donor for the introduction of specific desirable traits (reviewed in Hajjar and Hodgkin, 2007), hybridization and recombination could also make quantitative impacts on plant performance, and the utilization of CWR may lead to long-term benefit from increasing genetic diversity of the crop selection pool. That is, a diversity-driven instead of phenotype-driven approach in selecting the candidate plant materials. Thus, this base-broadening pre-breeding aims to develop germplasm ready to go under a changing climate or epidemic evolution. So, other than the conservation of CWR species, the availability of useful genetic information for breeders, as well as the ability to access existing diversity are other major tasks. A systematic high-throughput platform of genotyping to help estimate the existing diversity in the genebank would be a goal to increase the efficiency of management.

7.2.2 Outcomes of LR-59 and LR-26

As RILs on average contain half of the wild genome, the chance of carrying deleterious genes in one individual was also higher. In addition, since there are multiple segments of wild genome in each chromosome in a RIL, using a RIL

population to identify one specific genetic effect is also more difficult comparing to using introgression line populations of which several generations of backcross have been carried on to narrow the introgression region to one segment per individual (Tanksley and Nelson, 1996; Tanksley et al., 1996).

However, RIL populations are very useful for the evaluation of complex traits; and therefore, are useful for a preliminary genetic and phenotypic study. In this study, interspecific RILs enabled me to explore the impacts of CWI on several complex traits of both agronomic and seed quality importance in a quantitative manner and help identify multiple genomic regions association to these quantitative variations. Furthermore, while the goal of this study was not to select for superior genotypes, the information may still help breeders to evaluate the chance to make a successful cross for the backcross population development. For instance, LR-59-81 (Figure 7.1, II) is an interspecific RIL that carries multiple disease resistances while also having an earlier emergence time compared to its Ler parent accession L01-827A (Figure 7.1, III,), a similar level of sucrose and TRFO concentration as the Lcu parent Eston (Figure 7.1, I), a moderate seed size, and nod-dehiscence pods. This RIL has been selected to develop several genetic populations to help understand the genetics of introgression.

Figure 7.1 The morphology of seeds among Eston (left, I), LR-59-81 (middle, II) and L01-827A (right, III).

7.2.3 Prospective future works and lentil introgression breeding

The germplasm and CWR conservation for lentil breeding was recognised in the early 20th century by another important Russian botanist, Elena Barulina (1895-1957). In lentil, genetic bottlenecks have threatened the genetic gain of lentil improvement globally (Erskine et al., 1994; Erskine et al., 1998), so to restore the genetic variability by expanding the genetic base is very important. Beyond local landraces and varieties, CWR are a potential resource of novel variability. Since lentil is a staple protein crop in many regions of the world, the conservation and utilization of CWR of lentil is a priority of global lentil pre-breeding project supported by Global Crop Diversity Trust (https://www.croptrust.org/crop/lentil/).

From studying the Lcu x Ler derived progeny, transgressive variations can be found from a wide-range of aspects with a hint that such a wide cross has potential for diversifying selection. For instance, other than the agronomic and seed quality variations characterized in this study, different responses to drought and light quality can be found in the wild parent accessions of LR-26 and LR-59 (Gorim and Vandenberg, 2017; Yuan et al., 2017) and among the Lens spp. core collection which suggests possible variation in the LR-59 and LR-26 RILs for other characteristics of plant adaptation. Furthermore, lentil is a staple source of protein and energy in many regions, especially the developing part of the world (Erskine et al., 2011), and a great source of prebiotic components and micronutrients (Johnson et al., 2013; Thavarajah et al., 2009; Thavarajah et al., 2011; Kumar et al., 2016). Thus, to discover the potential of introgression breeding for improving nutritional quality is another highly desirable goal. For the future, to investigate segregation of additional traits using LR-26 and LR-59, as well as to screen the wide Lens collection for other potential genetic resources, are important goals.

From this study, a cross of primary x tertiary genepool combination can be very problematic in many aspects because of their evolutionary genomic divergence.

Although the genepool concept (Harlan and de Wet, 1971) help classify the availability of genetic resources of crop species, there is still a need for a more specific classification to identify the diversity between candidate germplasm and cultivars. Therefore, exploiting modern genomic technology in a diversity panel of available genetic resources is very important in foreseeing the values and challenges for planning a better breeding strategy. For instance, a genetic analysis such as a genome-wide association study on the core collection of *Lens* spp. may help identify superior alleles and identify high potential parental lines for the development of other interspecific or intraspecific breeding populations. As whole-genome re-sequencing technology becomes more available, scientists will also have the chance to assess haplotype diversity and to detect structural variations among genomes to assess the genetic divergence among diverse materials and to select easier-to-handle genetic resources. Moreover, the development of more interspecific and intraspecific populations is an important goal now. Increasing the diversity in the selection pool may increase the chance to select for environment-ready and epidemic-ready germplasm for a sustainable lentil improvement. Lastly, the construction of genetic maps of multiple interspecific and intraspecific populations can help in the design of molecular markers for further marker-assisted backcross selection and gene pyramiding purposes. These approaches will not only enable breeders to view the impacts and utility of CWI in a quantitative level, but also allow them to develop more efficient breeding strategies.

REFERENCES

Abbo, S., Ladizinsky, G., and Weeden, N. F. (1991). Genetic analysis and linkage study of seed weight in lentil. Euphytica 58:259-266.

Agil, R., Gaget, A., Gliwa, J., Avis, T. J., Willmore, W. G., and Hosseinian, F. (2013). Lentils enhance probiotic growth in yogurt and provide added benefit of antioxidant protection. LWT Food Sci. Technol. 50:45-49.

Amarowicz, R., and Pegg, R. B. (2008). Legumes as a source of natural antioxidants. Eur. J. Lipid Sci. Tech. 110:865-878.

Ballester, P., and Ferrándiz, C. (2017). Shattering fruits: variations on a dehiscent theme. Curr. Opin. Plant Biol. 35:68-75.

Barulina, H. (1930). Lentils of the USSR and other countries. Suppl. 40 to the Bulletin of Applied Botany, Genetics and Plant Breeding, 265-304. Leningrad (English Summary).

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., and Tanksley, S. (1998). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. Theor. Appl. Genet. 97:381-397.

Bessey, C. E. (1906). Crop improvement by utilizing wild species. J. Hered. 2:112-118.

Bett, K. E., Ramsay, L., Sharpe, A. G., Cook, D. R., Penmetsa, R. V., Stonehouse, R., Wong, M., Chan, C., Vandenberg, A., VanDeynze, A., Coyne, C.J ., McGee, R., Main, D., Dolezel, J., Edwards, D., Kaur, S., Udupa, S. M., and Kumar, S. (2014). Lentil genome sequencing: establishing a comprehensive platform for molecular breeding," in Proceedings of International Food Legumes Research Conference (IFLRC-VI) and ICCLG-VII (Saskatoon, SK: Crop Development Center) 19.

Bhadauria, V., Ramsay, L., Bett, K.E., and Banniza, S. (2017). QTL mapping reveals genetic determinants of fungal disease resistance in the wild lentil species *Lens ervoides*. Sci. Rep. 7:3231.

Bhatty, R. S. (1984). Relationship between physical and chemical characters and cooking quality in lentil. J. Agric. Food Chem. 32:1161-1166.

Bicer, B. T., and Sakar, D. (2004). Genetic variability and heritability for grain yield and other characters in lentil. J. Biol. Sci. 4: 216-218.

Bliss, F. A., Arulsekar, S., Foolad, M. R., Becerra, V., Gillen, A. M., Warburton, M. L., Dandekar, A. M., Kocsisne, G. M., and Mydin, K. K. (2002). An expanded genetic linkage map of Prunus based on an interspecific cross between almond and peach. Genome 45: 520-529.

Blöchl, A., Peterbauer, T., and Richter, A. (2007). Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. Plant Physiol. 164:1093-1096.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Brown, A. H. D., Lawrence, G. J., Jenkin, M., Douglass, J., and Gregory, E. (1989). Linkage drag in backcross breeding in barley. J. Hered. 80:234-239.

Buchwaldt, L., Anderson. K. L., Morrall. R. A. A., Gossen. B. D., and Bernier, C. C. (2004). Identification of lentil germplasm resistant to *Colletotrichum truncatum* and characterization of two pathogen races. Phytopathology 94:236-249.

Cameron, J. N., Han, Y., Wang, L., and Beavis, W. D. (2017). Systematic design for trait introgression projects. Theor. Appl. Genet. 130:1993-2004.

Campos-Vega, R., Loarca-Piña, G., and Oomah, B. D. (2010). Minor components of pulses and their potential impact on human health. Food Res. Int. 43:461-482.

Cannon, E. K. S., and Cannon, S. B (2011). Chromosome visualization tool: a whole genome viewer. Int. J. Plant Genom. 2011:373875.

Carew, R., Florkowski, W. J., and Zhang, Y. (2013). Review: Industry levy-funded pulse crop research in Canada: Evidence from the Prairie Provinces. Can. J. Plant Sci. 93:1017-1028.

Chowdhury, M. A., Andrahennadi, C. P., Slinkard, A. E., and Vandenberg, A. (2001). RAPD and SCAR markers for resistance to ascochyta blight in lentil. Euphytica 118: 331-337.
Cohen, D., Ladizinsky, G., Ziv, M., Muehlbauer, F. J. (1984). Rescue of interspecific lens hybrids by means of embryo culture. Plant Cell Tissue Organ. Cult. 3:343-347.
Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196.
Costa, G. E., Queiroz-Monici, K. D., Reis, S. M., and Oliveira, A. C. (2006). Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. Food Chem. 94: 327-330.
Cubero, J. I., Pérez de la Vega, M., and Fratini, R. (2009). Origin, phylogeny, domestication and spread. In: W. Erskine, F. J. Muehlbauer, A. Sarker, and B. Sharma (Eds.), The lentil: botany, production and uses (pp. 13-33). Wallinford, UK: CAB International.
Dante, R. A., Larkins, B. A., and Sabelli, P. A. (2014). Cell cycle control and seed development. Front. Plant Sci. 5:493.
da S. Queiroz-Monici, K., Costa, G. E., da Silva, N., Reis S. M., and de Oliveira, A. C. (2005). Bifidogenic effect of dietary fiber and resistant starch from leguminous on the intestinal microbiota of rats. Nutrition 21:602-608.
Davies, J. F. (1945). The effect of some environmental factors on the set of pods and yield of white pea beans. Acad. J. Agric. Res. 70:237-249.
Delouche, J. C. (1980). Environmental effects on seed development and seed quality. HortScience 15:775-780.
deMan, L., deMan, J., and Buzzell, R. (1987). Composition and properties of soymilk and tofu made from Ontario light hilum soybeans. J. Inst. Can. Science et Technol. Aliment. 20:363-367.
Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. Crop Sci. 57:1070-1082.
DellaValle, D. M., Thavarajah, D., Thavarajah, P., Vandenberg, A., and Glahn, R. P. (2013). Lentil (*Lens culinaris* L.) as a candidate crop for iron biofortification: Is there genetic potential for iron bioavailability? Field Crop. Res. 2013, 144, 119–125.
De Vicente, M. C., and Tanksley, S. D. (1993). QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585-596.
Dierkinga, E. C., and Bilyeu, K. D. (2009). Raffinose and stachyose metabolism are not required for efficient soybean seed germination. Plant Physiol. 166:1329-1335.
Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. 127:1309-1321.
Dong, Y., Yang, X., Liu, J., Wang, B.-H., Liu, B.-L., and Wang, Y.-Z. (2014). Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. Nat. Commun. 5:3352.
Dong, Y., and Wang, Y.-Z. (2015). Seed shattering: from models to crops. Front. Plant Sci. 24:476.
Erskine, W. (1985). Selection for pod retention and pod indehiscence in lentils. Euphytica 34:105-112.
Erskine, W., Williams, P. C., and Nakkoul, H. (1985). Genetic and environmental variation in the seed size, protein, yield, and cooking quality of lentils. Field Crop. Res. 12:153-161.
Erskine, W., Adham, Y., and Holly, L. (1989). Geographic distribution of variation in quantitative traits in a world lentil collection. Euphytica 43:97-103.
Erskine, W., Tufail, M., Russell, A., Tyagi, M. C., Rahman, M. M., and Saxena, M. C. (1994). Current and future strategies in breeding lentil for resistance to biotic and abiotic stresses. Euphytica 73: 127-135.
Erskine, W., Sarker, A., and Kumar, S. (2011). Crops that feed the world 3. Investing in lentil improvement toward a food secure world. Food Sec. 3:127.
Eujayl, I., Baum, M., Powell, W., Erskine, W., and Pehu, E. (1998). A genetic linkage map of lentil (*Lens* sp.) based on RAPD and AFLP markers using recombinant inbred lines. Theor. Appl. Genet. 97:83-89.

Food and Agriculture Organization of the United Nations. (2016). FAOSTAT Database. Rome, Italy: FAO. Retrieved May 3, 2018 from http://www.fao.org/faostat/en/#data/QC

Fedoruk, M. J., Vandenberg, A., and Bett, K. E. (2013) Quantitative trait loci analysis of seed quality characteristics in lentil using single nucleotide polymorphism markers. Plant Genom. 6:3.

Fernández-Aparicio, M., Sillero, J. C., and Rubiales, D. (2009). Resistance to broomrape in wild lentils (*Lens* spp.). Plant Breed. 128:266-270.

Fernie, A. R., Tadmor, Y., and Zamir, D. (2006). Natural genetic variation for improving crop quality. Curr. Opin. Plant Biol. 9:196-202.

Fiala, J. V., Tullu, A., Banniza, S., Se´guin-Swartz, G., and Vandenberg, A. (2009). Interspecies transfer of resistance to anthracnose in lentil (*Lens culinaris* Medik.). Crop Sci. 49:825-830.

Fleming, S. E. (1982). Flatulence activity of the smooth-seeded field pea as indicated by hydrogen production in the rat. J. Food Sci. 47:12-15.

Ford, R., Pang, E. C. K., and Taylor, P. W. J. (1999). Genetics of resistance to ascochyta blight (*Ascochyta lentis*) of lentil and the identification of closely linked RAPD markers. Theor. Appl. Genet. 98:93-98.

Fratini, R., Durá, Y., García, P., and Pérez De La Vega, M. (2007). Identification of quantitative trait loci (QTL) for plant structure, growth habit and yield in lentil. Span. J. Agric. Res. 5:348–356.

Fratini, R., and Ruiz, M. L. (2010). Wide crossing in lentil through embryo rescue. In: T. A. Thorpe, and E. C. Yeung (Eds.), *Plant embryo culture: Methods and protocols* (pp. 131-139). New York, NY: Springer.

Frias, J., Diaz-Pollan, C., Hedley, C. L., and Vidal-Valverde, C. (1996). Evolution and kinetics of monosaccharides, disaccharides and a-galactosides during germination of lentils. Eur. Food Res. Technol. 202:3539.

Frias, J., Bakhsh, A., Jones, D. A., Arthur, A. E., Vidal-Valverde, C., Rhodes, M. J. C., and Hedley, C. L. (1999). Genetic analysis of the raffinose oligosaccharide pathway in lentil seeds. J. Exp. Bot. 50:469-476.

Frischa, M., Bohna, M., and Melchinger, A. E. (1998). Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci. 39:1295-1301.

Fuller, D. Q. (2007). Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the old world. Ann. Bot. 100:903-924.

Gangl, R., and Tenhaken, R. (2016). Raffinose family oligosaccharides act as galactose stores in seeds and are required for rapid germination of Arabidopsis in the dark. Front Plant Sci. 7: 1115.

Gangola, M. P., Jaiswal, S., Khedikar, Y. P., and Chibbar, R. N. (2014). A reliable and rapid method for soluble sugars and RFO analysis in chickpea using HPAEC–PAD and its comparison with HPLC–RI. Food Chem. 154:127-133.

Gangola, M. P., Jaiswal, S., Kannan, U., Gaur, P. M., Båga, M., and Chibbar, R. N. (2016). Galactinol synthase enzyme activity influences raffinose family oligosaccharides (RFO) accumulation in developing chickpea (*Cicer arietinum* L.) seeds. Phytochemistry, 125, 88-98. doi:10.1016/j.phytochem.2016.02.009

Gasim, S., Abel, S., and Link, W. (2004). Extent, variation and breeding impact of natural cross-fertilization in German winter faba beans using hilum colour as marker. Euphytica 136:193-200.

Ghavidel, R. A., and Prakash, J. (2007). The impact of germination and dehulling on nutrients, antinutrients, in vitro iron and calcium bioavailability and in vitro starch and protein digestibility of some legume seeds. Lebenson. Wiss. Technol. 40:1292-1299

Gorim, L. Y., and Vandenberg, A. (2017). Evaluation of wild lentil species as genetic resources to improve drought tolerance in cultivated lentil. Front. Plant Sci. 8:1129.

Gujaria-Verma, N., Vail, S. L., Carrasquilla-Garcia, N., Penmetsa, R.V., Cook, D. R., Farmer, A. D., Vandenberg, A., and Bett, K. E. (2014). Genetic mapping of legume orthologs reveals high conservation of synteny between lentil species and the sequenced genomes of Medicago and chickpea. Front. Plant. Sci. 5:676.

Gur, A., and Zamir, D. (2004). Unused natural genetic variation can lift yield barrier in plant breeding. Plos Biol. 2:e245.

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. Euphytica 156:1-13.

Hamdi, A., Erskine, W., and Gates, P. (1991). Relationships among economic characters in lentil. Euphytica 57:109-116.
Harlan, J. R., and de Wet, J. M. J. (1971). Toward a rational classification of cultivated plants. Taxon 20:509-517.
Harlan, J. R. (1976). Genetic resources in wild relatives of crops. Crop Sci. 16:329-333.
Hefnawy, T. H. (2011). Effect of processing methods on nutritional composition and anti-nutritional factors in lentils (*Lens culinaris*) Ann. Agric. Sci. 56:57-61.
Hitz, W. D., Carlson, T. J., Kerr, P. S., and Sebastian, S. A. (2002). Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiol. 128:650-660.
Hoffman, D. L., Muehlbauer, F. J., and Ladizinsky, G. (1988). Morphological Variation in *Lens* (Leguminosae). System. Bot. 13:87-96.
Horbowicz, M., and Obendorf, R. L. (1994). Seed desiccation tolerance and stability: dependence on flatulence producing oligosaccharide and cyclitols-review and survey. Seed Science Research 4:385-405.
Idrissi, O., Udupa, S. M., De Keyser, E., McGee, R. J., Coyne, C. J., Saha, G. C., Muehlbauer, F. J., Van Damme, P., and De Riek, J. (2016). Identification of quantitative trait loci controlling root and shoot traits associated with drought tolerance in a lentil (*Lens culinaris* Medik.) recombinant inbred line population. Front. Plant Sci. 7:1174.
Imai, I., Kimball, J. A., Conway, B., Yeater, K. M., McCouch, S. R., and McClung, A. (2013). Validation of yield-enhancing quantitative trait loci from a low-yielding wild ancestor of rice. Mol. Breed. 32:101-120.
Jain, N., Dikshit, H. K., Singh, D., Singh, A., and Kumar, H. (2013). Discovery of EST-derived microsatellite primers in the legume Lens culinaris (Fabaceae). Appl. Plant Sci. 1:1200539.
James, W., and Major, A. (2002). Pulses and lipaemia, short- and long-term effect: potential in the prevention of cardiovascular disease. Br. J. Nutr. 88: S263-S271.
Jansen, R. C. (1993). Interval mapping of multiple quantitative trait loci. Genetics 135:205-211.
Johnson, C. R., Thavarajah, D., Combs Jr., G. F., and Thavarajah, P. (2013). Lentil (*Lens culinaris* L.): A prebiotic-rich whole food legume. Food Res. Int. 51:107-113.
Jones, D. A., Dupont, M. S., Ambrose, M. J., Frias, J., and Hedley, C. L. (1999). The discovery of compositional variation for the raffinose family oligosaccharides in pea seeds. Seed Sci. Res. 9:305-310.
Joshi, M., Aldred, P., McKnight, S., Panozzo, J. F., Kasapis, S., Adhikari, R., and Adhikari, B. (2012). Physicochemical and functional characteristics of lentil starch. Carbohydr. Polym. 92:1484-1496.
Kant, P., Materne, M., Rodda, M. S., and Slater, A. T. (2017) Screening lentil germplasm for stemphylium blight resistance. Australas. Plant Pathol. 46:129-136.
Kahraman, A., Kusmenoglu, I., Aydin, N., Aydogan, A., Erskine, W., and Muehlbauer, F. J. (2004). QTL mapping of winter hardiness genes in lentil. Crop Sci. 44:13-22.
Kaur, S., Cogan, N. O., Pembleton, L. W., Shinozuka, M., Savin, K. W., Materne, M., and Foster, J. W. (2011). Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genom. 12:265.
Kaur, S., Cogan, N. O., Stephens, A., Noy, D., Butsch, M., Forster, J. W., and Materne, M. (2014). EST-SNP discovery and dense genetic mapping in lentil (Lens culinaris Medik.) enable candidate gene selection for boron tolerance. Theor. Appl. Genet. 127:703-713.
Khazaei, H., Caron, C. T., Fedoruk, M., Diapari, M., Vandenberg, A., Coyne, C. J., McGee, R., and Bett, K. E. (2016). Genetic diversity of cultivated lentil (*Lens culinaris* Medik.) and its relation to the world's agro-ecological zones. Front Plant Sci. 7:1093.
Khazaei, H., Fedoruk, M., Caron, C. T., Vandenberg, A., and Bett, K. E. (2017). Single nucleotide polymorphism markers associated with seed quality characteristics of cultivated lentil. Plant Genom. 11:170051.
Kim, S. J., de Souza, R. J., Choo, V. L., Ha, V., Cozma, A. I., Chiavaroli, L., Mirrahimi, A., Mejia, S. B., Di Buono, M., Bernstein, A. M., Leiter, L. A., Kris-Etherton, P. M., Vuksan, V., Beyene, J., Kendall, C. W. C., Jenkins,

D. J. A., and Sievenpiper, J. L. (2016). Effects of dietary pulse consumption on body weight: a systematic review and meta-analysis of randomized controlled trials. Am. J. Clin. Nutr. 103:1213-1223.
Koster K. L., and Leopold, A. C. (1988). Sugar and desiccation tolerance in seeds. Plant Physiology 88:829-832.
Kumar, S., Rajendran, K., Kumar, J., Hamwieh, A., and Baum, M. (2015). Current knowledge in lentil genomics and its application for crop improvement. Front. Plant Sci. 6:78.
Kumar, J., Gupta, D. S., Kumar, S., Gupta, S., and Singh, N. P. (2016). Current knowledge on genetic biofortification in lentil. J. Agric. Food Chem. 64:6383-6396.
Kuo, T. M., Van Middlesworth, J. F., and Wolf, W. J. (1988). Content of raffinose oligosaccharides and sucrose in various plant seeds. J. Agric. Food Chem. 36:32-36.
Ky, C. L., Barre, P., Lorieux, M., Trouslot, P., Akaffou, S., Louarn, J., Charrier, A., Hamon, S., and Noirot, M. (2000). Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.) Theor. Appl. Genet. 101:669-676.
Ladizinsky, G. (1979). The genetics of several morphological traits in the lentil. J. Hered. 70:135-137.
Ladizinsky, G. (1985). The genetics of hard seed coat in the genus *Lens*. Euphytica 34:539-543.
Ladizinsky, G., Cohen, D., and Muehlbauer, F. J. (1985). Hybridization in the genus *Lens* by means of embryo culture. Theor. Appl. Genet. 70:97-101.
Lande, R., and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics. 124:743-756.
Lander, E. S., and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121:185-199.
Laserna-Ruiz, I., De-Los-Mozos-Pascual, M., Santana-Méridas, O., Sánchez-Vioque, R., and Rodríguez-Conde, M. F. (2012). Screening and selection of lentil (*Lens* Miller) germplasm resistant to seed bruchids (*Bruchus* spp.). Euphytica 188:153-162.
Leterme, P. (2002). Recommendations by health organizations for pulse consumption. Br. J. Nutr. 88: S239-S242.
Li, T., Zhang, Y., Wang, D., Liu, Y., Dirk, L. M. A., Goodman, J., Downie, A. B., Wang, J., Wang, G., and Zhao, T. (2017). Regulation of seed vigor by manipulation of raffinose family oligosaccharides in maize and *Arabidopsis thaliana*. Mol. Plant 10:1540-1555.
Lin, Z., Li, X., Shannon, L. M., Yeh, C. T., Wang, M. L., Bai, G., Peng, Z., Li, J., Trick, H. N., Clemente, T. E., Doebley, J., Schnable, P. S., Tuinstra, M. R., Tesso, T. T., White, F., and Yu, J. (2012). Parallel domestication of the Shattering1 genes in cereals. Nat. Genet. 44:720-724.
Lyttle, T. W. (1991). Segregation distorters. Annu. Rev. Genet. 25:511-557.
Martin, G., Carreel, F., Coriton, O., Hervouet, C., Cardi, C., Derouault, P., Roques, D., Salmon, F., Rouard, M., Sardos, J., Labadie, K., Baurens, F.-C., and D'Hont, A. (2017). Evolution of the banana genome (*Musa acuminata*) is impacted by large chromosomal translocations. Mol. Biol. Evol. 34:2140–2152.
Martínez-Villaluenga, C., Frias, J., and Vidal-Valverde, C. (2008). Alpha-galactosides: Antinutritional factors or functional ingredients? Crit. Rev. Food Sci. Nutr.48:301-316.
McCouch, S. R., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septiningsih, E., Edwards, J., Moncada, P., Xiao, J., Garris, A., Tai, T., Martinez, C., Tohme, J., Sugiono, M., McClung, A., Yuan, L. P., and Ahn, S.-N. (2007). Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. Euphytica 154:317-339.
McCrory, M. A., Hamaker, B. R., Lovejoy, J. C., Eichelsdoerfer, P. E. (2010). Pulse consumption, satiety, and weight management. Adv. Nutr. 1:17-30.
Muzquiz, M., Varela. A., Burbano. C., Cuadrado. C., Guillamo´n, E., and Pedrosa, M. M. (2012). Bioactive compounds in legumes: pronutritive and antinutritive actions. Implications for nutrition and health. Phytochem. Rev. 11:227-244.
Oomah, B. D., Caspar, F., Malcolmson, L. J., and Bellido, A. -S. (2011). Phenolics and antioxidant activity of lentil and pea hulls. Food Res. Int. 44:436-441.

Østerberg, J. T., Xiang, W, Olsen, L. I., Edenbrandt, A. K., Vedel, S. E., Christiansen, A., Landes, X., Andersen, M. M., Pagh, P., Sandøe, P., Nielsen, J., Christensen, S. B., Thorsen, B. J., Kappel, K., Gamborg, C., and Palmgren, M. (2017). Accelerating the domestication of new crops: Feasibility and approaches. Trends Plant Sci. 22: 373-384.

Peterbauer, T., and Richter, A. (2001). Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Sci. Res. 11:185-197.

Placido, D. F., Campbell, M. T., Folsom, J. J., Cui, X., Kruger, G. R., Baenziger, P. S., and Walia, H. (2013). Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. Plant Physio. 161:1806-1819.

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92-102.

Podder, R. (2012). The potential for breeding for stemphylium blight resistance in the genus Lens (Master's thesis, University of Saskatchewan, Saskatoon, Canada). Retrieved from https://ecommons.usask.ca/handle/10388/ETD-2012-10-739.

Podder, R., Banniza, S., and Vandenberg, A. (2013). Screening of wild and cultivated lentil germplasm for resistance to Stemphylium blight. Plant Genet. Resour. 11:26-35.

Podder, R., Tar'an, B., Tyler, R. T., Henry, C. J., DellaValle, D. M., and Vandenberg, A. (2017). Iron fortification of lentil (*Lens culinaris* Medik.) to address iron deficiency. Nutrients 9: 863.

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Díez, M. J., Fita, A., Herraiz, F. J., Rodríguez-Burruezo, A., Soler, S., Knapp, S., and Vilanova, S. (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. Euphytica 213:158.

Queiroz-Monici, K. da S., Costaa, G. E. A., da Silva, N., Reis, S. M. P. M., and de Oliveira, A. C. (2005). Bifidogenic effect of dietary fiber and resistant starch from leguminous on the intestinal microbiota of rats. Nutrition 21:602-608.

Rackis, J. J. (1981). Flatulence caused by soy and its control. J. Amer. Oil Chem. Soc., 58:503-509.

Rick, C. M., and Smith, P. G. (1953). Novel variation in tomato species hybrids. Am. Nat. 87:359-373.

Rick, C. (1969). Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum*: segregation and recombination. Genetics 62:753-768.

Roberts, E. H., Summerfield, R. J., Muehlbauer, F. J., and Short, R. W. (1986). Flowering in lentil (Lens culinaris Medic.): the duration of the photoperiodic inductive phase as a function of accumulated daylength above the critical photoperiod. Ann. Bot. 58:235-248.

Roy, F., Boye, J. I., and Simpson, B. K. (2010) Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. Food Res. Int. 43:432-442.

Rubeena, Taylor, P. W. J., Ades, P. K., and Ford, R. (2006). QTL mapping of resistance in lentil (*Lens culinaris* ssp. *culinaris*) to ascochyta blight (*Ascochyta lentis*). Plant Breed. 125:506-512.

Saha, G. C., Sarker, A., Chen, W., Vandemark, G. J., and Muehlbauer, F. J. (2009). Inheritance and linkage map positions of genes conferring resistance to stemphylium blight in lentil. Crop Sci. 50:1831-1839.

Salon, C., Munier-Jolain, N., Duc, G., Voisin, A.-S., Grandgirard, D., Larmure, A., Emery, R., and Ney, B. (2001). Grain legume seed filling in relation to nitrogen acquisition: A review and prospects with particular reference to pea. Agronomie 21:539-552.

Sengupta, S., Mukherjee, S., Basak, P., and Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. Front. Plant Sci. 6: 656.

Servin, B., Martin, O. C., Mézard, M., Hospital, F. (2004). Toward a theory of marker-assisted gene pyramiding. Genetics 168:513-523.

Sharma, B., and Emami, M. K. (2002). Discovery of a new gene causing dark green cotyledons and pathway of pigment synthesis in lentil (*Lens culinaris* Medik) Euphytica 124: 349-353.

Sharma, S., Upadhyaya, H. D., Varshney, R. K., and Gowda, C. L. L. (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. Front. Plant Sci. 4: 309.

Sharpe, A. G., Ramsay, L., Sanderson, L.-A., Fedoruk, M. J., Clarke, W. E., Li, R., Kagale, S., Vijayan, P., Vandenberg, A., and Bett, K. E. (2013). Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. BMC Genom. 14:192.

Singh, M., Rana, M. K., Kumar, K., Bisht, I. S., Dutta, M., Gautam, N. K., Sarker, A., and Bansal, K. C. (2013) Broadening the genetic base of lentil cultivars through inter-sub-specific and interspecific crosses of *Lens* taxa. Plant Breed. 132:667-675.

Singh, M., Bisht, I. S., Kumar, S., Dutta, M., and Chander, K. (2014). Global wild annual *Lens* collection: a potential resource for lentil genetic base broadening and yield enhancement. Plos ONE 9: e107781.

Singh, M., Rana, J. C., Singh, B, Kumar, S., Saxena, D.R., Saxena, A., Rizvi, A. H., and Ashutosh, S. (2017a). Comparative agronomic performance and reaction to Fusarium wilt of *Lens culinaris* × *L. orientalis* and *L. culinaris* × *L. ervoides* derivatives. Front. Plant Sci. 8:1162.

Singh, D., Singh, C. K., Taunk, J., Tomar, R. S. S., Chaturvedi, A. K., Gaikwad, K., and Pal, M. (2017b). Transcriptome analysis of lentil (*Lens culinaris* Medikus) in response to seedling drought stress. BMC Genom. 18:206.

Slavin, J. (2013). Fiber and prebiotics: Mechanisms and health benefits. Nutrients 5:1417-1435.

Soller, M., Brody, T., and Genizi, A. (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47:35-39.

Summerfield, R. J., Roberts, E. H., Erskine, W., and Ellis, R. H. (1985) Effects of temperature and photoperiod on flowering in lentils (*Lens culinaris* Medic.). Ann. Bot. 56:659-671.

Suryapani, S., Umar, S., Malik, A. A., and Ahmad, A. (2012). Symbiotic nitrogen fixation by lentil improves biochemical characteristics and yield of intercropped wheat under low fertilizer input. J. Crop Improv. 27:53-66.

Tadmor, Y., Zamir, D., and Ladizinsky, G. (1987). Genetic mapping of an ancient translocation in the genus *Lens*. Theor. Appl. Genet. 73:883-892.

Tadmor, Y., Fridman, E., Gur, A., Larkov, O., Lastochkin, E., Ravid, U., Zamir, D., and Lewinsohn, E. (2002). Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. J. Agric. Food Chem. 50:2005-2009.

Tahir, M., Vandenberg, A., and Chibbar, R. (2011a). Influence of environment on seed soluble carbohydrates in selected lentil cultivars. J. Food Comp. Anal. 24:596-602.

Tahir, M., Lindeboom, N., Båga, M., Vandenberg, A., and Chibbar, R. (2011b). Composition and correlation between major seed constituents in selected lentil (*Lens culinaris*. Medik) genotypes. Can. J. Plant Sci. 91:825-835.

Tahir, M., Båga, M., Vandenberg, A., and Chibbar, R. (2012). An assessment of raffinose family oligosaccharides and sucrose concentration in genus *Lens*. Crop Sci. 52:1713-1720.

Tambal, H. A. A., Erskine, W., Baalbaki, R., and Zaiter, H. (2000). Relationship of flower and pod numbers per inflorescence with seed yield in lentil. Expl. Agric. 36:369-378.

Tanksley, S. D., Young, N. D., Paterson, A. H., and Bonierbale, M. W. (1989). RFLP mapping in plant breeding: New tools for an old science. Nat. Biotechnol. 7:257-264.

Tanksley, S. D., and Nelson, J. C. (1996). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92:191-203.

Tanksley, S., and McCouch, S. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063-1066.

Tar'an, B., Buchwaldt, L., Tullu, A., Banniza, S., Warkentin, T. D., and Vandenberg, A. (2003). Using molecular markers to pyramid genes for resistance to ascochyta blight and anthracnose in lentil (*Lens culinaris* Medik). Euphytica 134:223-230.

Thavarajah, D., Thavarajah, P., Sarker, A., and Vandenberg, A. (2009). Lentils (*Lens culinaris* Medikus Subspecies *culinaris*): A whole food for increased iron and zinc Intake. J. Agric. Food Chem. 57:5413-5419.

Thavarajah, D., Thavarajah, P., Wejesuriya, A., Rutzke, M., Glahn, R. P., Combs, G. F., and Vandenberg, A. (2011). The potential of lentil (*Lens culinaris* L.) as a whole food for increased selenium, iron, and zinc intake: Preliminary results from a 3 year study. Euphytica 180:123-128.

Thormann, I., Parra-Quijano, M., and Endresen, D. (2014) *Predictive characterization of crop wild relatives and landraces: Technical guidelines version 1*. Rome, Italy: Biodiversity International.

Topping, D. L., and Clifton, P. M. (2001). Short-chain fatty acids and human colonic function: Roles of resistant starch and nonstarch polysaccharides. Physiol. Rev. 81:1031-1064.

Tullu, A., Kusmenoglu, I., McPhee, K. E., and Muehlbauer, F. J. (2001) Characterization of core collection of lentil germplasm for phenology, morphology, seed and straw yields. Genet. Resour. Crop Ev. 48:143-152.

Tullu, A., Buchwaldt, L., Lulsdorf, M., Banniza, S., Barlow, B., Slinkard, A. E., Sarker, A., Tar'an, B., Warkentin, T., and Vandenberg, A. (2006). Sources of resistance to anthracnose (*Colletotrichum truncatum*) in wild *Lens* species. Genet. Resour. 53:111-119.

Tullu, A., Tar'an, B., Warkentin, T., and Vandenberg, A. (2007). Construction of an intraspecific linkage map and QTL analysis for earliness and plant height in lentil. Crop Sci. 48:2254-2264.

Tullu, A., Bett, K., Banniza, S., Vail, S., and Vandenberg, A. (2013). Widening the genetic base of cultivated lentil through hybridization of *Lens culinaris* 'Eston' and *L. ervoides* accession IG 72815. Can. J. Plant Sci. 93:1037-1047.

Tullu, A., Banniza, S., Tar'an, B., Warkentin, T., and Vandenberg, A. (2010). Sources of resistance to ascochyta blight in wild species of lentil (*Lens culinaris* Medik). Genet. Resour. Crop Evol. 57:1053-1063.

Vail, S., and Vandenberg, A. (2010). Evaluation of a clonal propagation protocol to obtain replicated disease data on infection by *Colletotrichum truncatum* in *Lens culinaris*. Crop Sci. 50:926-932.

Vail, S., and Vandenberg, A. (2011). Genetic control of interspecific-derived and juvenile resistance in lentil to *Colletotrichum truncatum*. Crop Sci. 51: 1481-1490.

Vail, S., Strelioff, J. V., Tullu, A., and Vandenberg, A. (2012). Field evaluation of resistance to *Colletotrichum truncatum* in *Lens culinaris*, *L. ervoides*, and *Lens ervoides* x *Lens culinaris* derivatives. Field Crops Res. 126: 145-151.

van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R., and Visser, B. (2009). Genetic erosion in crops: concept, research results and challenges. Plant Genet. Resour. C. 8:1-15.

Varshney, R. K., Ribaut, J.-M., Buckler, E. S., Tuberosa, R., Rafalski, J. A., and Langridge, P. (2012). Can genomics boost productivity of orphan crops? Nat. Biotech. 30:1172-1175.

van Oss, H., Aron, Y., and Ladizinsky, G. (1997). Chloroplast DNA variation and evolution in the genus Lens Mill. Theor. Appl. Genet. 94: 452-457.

Vidal-Valverde, C., and Frias, J. (1992). Changes in carbohydrates during germination of lentil. Z. Lebensm. Unters. For. 194: 46464.

Vidal-Valverde, C., Frias, J., Prodanov, M., Tabera, J., Ruiz, R., and Bacon, J. (1993). Effect of natural fermentation on carbohydrates, riboflavin and trypsin inhibitor activity of lentils. Eur. Food Res. Technol. 197:449-452.

Wang, T. L., Domoney, C., Hedley, C. L., Casey, R., and Grusak, M. A. (2003). Can we improve the nutritional quality of legume seeds? Plant Physiol. 131:886-891

Wang, N., Hatcher, D. W., Toews, R., and Gawalko, E. J. (2009). Influence of cooking and dehulling on nutritional composition of several varieties of lentils (*Lens culinaris*). LWT Food Sci. Technol. 42:842-848

Weeden, N. F. (2007). Genetic changes accompanying the domestication of *Pisum sativum*: is there a common genetic basis to the 'domestication syndrome' for legumes? Ann Bot. 100:1017-25.

Weiss, E., and Zohary, D. (2011). The Neolithic Southwest Asian founder crops. Current Anthropology, 52:S237-S254.

Wickland, D. P., Battu, G., Hudson, K. A., Diers, B. W., and Hudson, M. E. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. BMC Bioinformatics 18:586.

Wilson, V. E., Law, A. G., and Warner, R. L. (1970). Inheritance of cotyledon colour in *Lens culinaris* Medik. Crop Sci. 10: 205-207.

Wilson, V. E., and Law, A. G. (1972). Natural crossing in *Lens esculenta* Moench. J. Am. Soc. Hortic. Sci. 97:142-143.

Wong, M. M. L., Gujaria-Verma, N., Ramsay, L., Yuan, H. Y., Caron, C., Diapari, M., Vandenberg, A., and Bett, K. E. (2015). Classification and characterization of species within the genus Lens using genotyping-by-sequencing (GBS). PLoS ONE 10: e0122025.

Xiao, J., Grandillo, S., Ahn, S. H., McCouch, S. R., Tanksley, S. D., Li, J., and Yuan, L. (1996). Genes from wild rice improve yield. Nature 384:223-224.

Yang, K., Ko, J.-M., Ha, T. J., Lee, Y.-H., Baek, I.-Y., Yang, T.-J., and Nou, I.-S. (2014). Development of molecular markers for low raffinose and stachyose in Korean soybean cultivars. Plant Breed. Biotechnol. 2:151-157.

Youngman, V. E. (1967). Lentils—a pulse of the Palouse. Econ. Bot. 22:135-139.

Yuan, H. Y., Saha, S., Vandenberg, A., and Bett, K. E. (2017) Flowering and growth responses of cultivated lentil and wild Lens germplasm toward the differences in red to far-red ratio and photosynthetically active radiation. Front. Plant Sci. 8:386.

Zamir, D., and Ladizinsky, G. (1984). Genetics of allozyme variants and linkage groups in lentil. Euphytica 33:329-336.

Zamir, D., and Tadmor, Y. (1986). Unequal segregation of nuclear genes in plants. Bot. Gaz. 147:355-358.

Zamir, D., Ekstein-Michelson, I., Zakay, Y., Navot, N., Zeidan, M., Sarfatti, M., Eshed, Y., Harel, E., Pleban, T., van-Oss, H., Kedar, N., Rabinowitch, H. D., and Czosnek, H. (1994). Mapping and introgression of a tomato yellow leaf curl virus tolerance gene, TY-1. Theor. Appl. Genet. 88:141-146.

Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. Nat. Rev. Genet. 2:983-989

Zamir, D. (2013). Where have all the crop phenotypes gone? PLOS Biol. 11:e1001595.

Zeng, Z. B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA. 90:10972-10976.

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

Zohary, D. (1972). The wild progenitor and the place of origin of the cultivated lentil: *Lens culinaris*. Econ. Bot. 26: 326-332.

Zohary, D. (2004). Unconscious selection and the evolution of domesticated plants. Econ. Bot. 58: 5-10.

APPENDICES

Appendix A

Figure A. The temperature and precipitation in CSSF13, CSSF14, CSSF15 and STH15 during the period of cultivation.

Appendix B

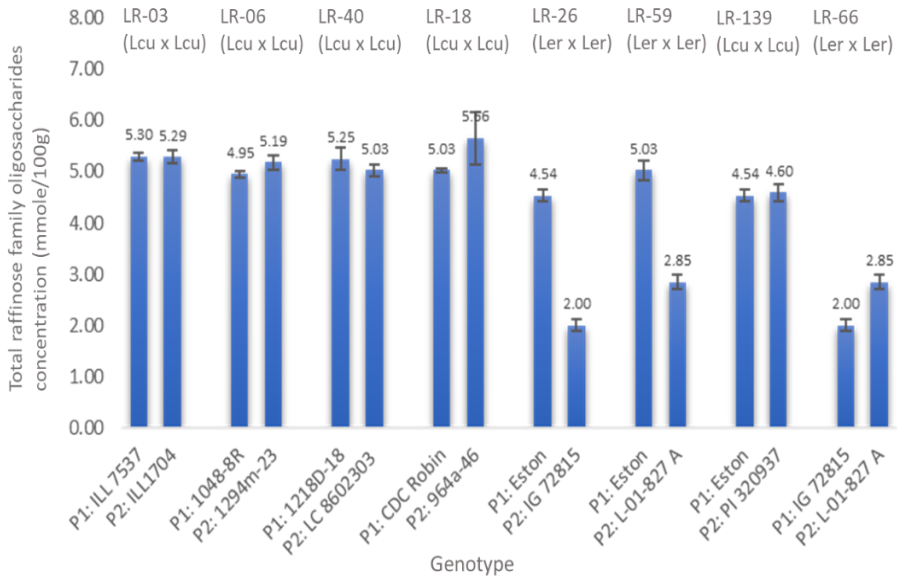


Figure B. Preliminary experiment to screen the TRFO content between the parents of the genetic populations generated by the CDC. Contrasting levels were found between the parents of LR-26 and

Appendix C

Table C.1: List of individual genotypes of LR-26, derived from Eston (*Lens culinaris*) x IG 72815 (*L. ervoides*) included in field trial.

Eston	LR-26-63	LR-26-118	LR-26-170	LR-26-220	LR-26-274
IG 72815	LR-26-64	LR-26-119	LR-26-171	LR-26-223	LR-26-275
LR-26-3	LR-26-65	LR-26-121	LR-26-172	LR-26-224	LR-26-276
LR-26-4	LR-26-66	LR-26-122	LR-26-173	LR-26-227	LR-26-280
LR-26-5	LR-26-67	LR-26-123	LR-26-175	LR-26-228	LR-26-281
LR-26-7	LR-26-70	LR-26-125	LR-26-176	LR-26-232	LR-26-282
LR-26-10	LR-26-75	LR-26-127	LR-26-178	LR-26-233	LR-26-283
LR-26-12	LR-26-77	LR-26-128	LR-26-180	LR-26-235	LR-26-288
LR-26-13	LR-26-78	LR-26-129	LR-26-181	LR-26-238	LR-26-290
LR-26-16	LR-26-79	LR-26-132	LR-26-182	LR-26-239	LR-26-292
LR-26-17	LR-26-83	LR-26-134	LR-26-183	LR-26-240	LR-26-293
LR-26-18	LR-26-84	LR-26-135	LR-26-184	LR-26-241	LR-26-294
LR-26-19	LR-26-85	LR-26-136	LR-26-186	LR-26-243	LR-26-296
LR-26-20	LR-26-86	LR-26-137	LR-26-187	LR-26-244	LR-26-297
LR-26-22	LR-26-87	LR-26-138	LR-26-188	LR-26-245	LR-26-298
LR-26-23	LR-26-90	LR-26-139	LR-26-193	LR-26-246	LR-26-299
LR-26-29	LR-26-91	LR-26-140	LR-26-194	LR-26-247	LR-26-300
LR-26-30	LR-26-95	LR-26-142	LR-26-196	LR-26-251	LR-26-301
LR-26-32	LR-26-98	LR-26-145	LR-26-198	LR-26-252	LR-26-303
LR-26-36	LR-26-99	LR-26-149	LR-26-200	LR-26-253	LR-26-304
LR-26-41	LR-26-105	LR-26-151	LR-26-202	LR-26-254	LR-26-306
LR-26-43	LR-26-107	LR-26-152	LR-26-203	LR-26-256	LR-26-307
LR-26-45	LR-26-108	LR-26-156	LR-26-204	LR-26-257	LR-26-311
LR-26-47	LR-26-110	LR-26-157	LR-26-205	LR-26-259	LR-26-312
LR-26-49	LR-26-111	LR-26-161	LR-26-206	LR-26-261	
LR-26-54	LR-26-112	LR-26-162	LR-26-209	LR-26-262	
LR-26-55	LR-26-113	LR-26-163	LR-26-210	LR-26-266	
LR-26-56	LR-26-115	LR-26-164	LR-26-215	LR-26-267	
LR-26-57	LR-26-116	LR-26-165	LR-26-216	LR-26-269	
LR-26-62	LR-26-117	LR-26-169	LR-26-219	LR-26-273	

Table C.2: List of individual genotypes of LR-59, derived from Eston (*Lens culinaris*) x L01-827A (*L. ervoides*) included in field trial.

Eston	LR-59-27	LR-59-55	LR-59-89	LR-59-122
L01-827A	LR-59-29	LR-59-56	LR-59-90	LR-59-126
LR-59-1	LR-59-30	LR-59-59	LR-59-91	LR-59-127
LR-59-2	LR-59-34	LR-59-60	LR-59-95	LR-59-128
LR-59-4	LR-59-35	LR-59-61	LR-59-96	LR-59-129
LR-59-5	LR-59-36	LR-59-62	LR-59-103	LR-59-130
LR-59-6	LR-59-37	LR-59-70	LR-59-104	LR-59-132
LR-59-7	LR-59-41	LR-59-71	LR-59-105	LR-59-133
LR-59-9	LR-59-42	LR-59-74	LR-59-106	
LR-59-10	LR-59-43	LR-59-76	LR-59-107	
LR-59-11	LR-59-44	LR-59-78	LR-59-112	
LR-59-14	LR-59-47	LR-59-80	LR-59-114	
LR-59-15	LR-59-49	LR-59-81	LR-59-115	
LR-59-23	LR-59-53	LR-59-86	LR-59-117	
LR-59-25	LR-59-54	LR-59-87	LR-59-121	

Table C.3: List of individual genotypes of LR-26, derived from Eston (*Lens culinaris*) x IG 72815 (*L. ervoides*) for linkage analysis and QTL mapping.

LR-26-3	LR-26-65	LR-26-125	LR-26-178	LR-26-233	LR-26-283
LR-26-4	LR-26-66	LR-26-127	LR-26-180	LR-26-235	LR-26-288
LR-26-5	LR-26-67	LR-26-128	LR-26-181	LR-26-238	LR-26-290
LR-26-7	LR-26-70	LR-26-129	LR-26-182	LR-26-239	LR-26-292
LR-26-10	LR-26-75	LR-26-132	LR-26-183	LR-26-240	LR-26-293
LR-26-12	LR-26-77	LR-26-134	LR-26-184	LR-26-241	LR-26-294
LR-26-13	LR-26-78	LR-26-135	LR-26-186	LR-26-243	LR-26-296
LR-26-16	LR-26-79	LR-26-136	LR-26-187	LR-26-244	LR-26-297
LR-26-17	LR-26-84	LR-26-137	LR-26-188	LR-26-245	LR-26-298
LR-26-18	LR-26-85	LR-26-138	LR-26-193	LR-26-246	LR-26-299
LR-26-19	LR-26-86	LR-26-139	LR-26-194	LR-26-247	LR-26-300
LR-26-20	LR-26-87	LR-26-140	LR-26-196	LR-26-251	LR-26-301
LR-26-22	LR-26-90	LR-26-142	LR-26-198	LR-26-252	LR-26-303
LR-26-23	LR-26-91	LR-26-145	LR-26-200	LR-26-253	LR-26-304
LR-26-29	LR-26-95	LR-26-149	LR-26-202	LR-26-254	LR-26-306
LR-26-30	LR-26-98	LR-26-151	LR-26-203	LR-26-256	LR-26-307
LR-26-32	LR-26-99	LR-26-156	LR-26-204	LR-26-257	LR-26-311
LR-26-36	LR-26-105	LR-26-157	LR-26-205	LR-26-259	
LR-26-41	LR-26-107	LR-26-161	LR-26-206	LR-26-261	
LR-26-43	LR-26-108	LR-26-162	LR-26-209	LR-26-262	
LR-26-45	LR-26-110	LR-26-163	LR-26-210	LR-26-266	
LR-26-47	LR-26-111	LR-26-164	LR-26-215	LR-26-267	
LR-26-49	LR-26-112	LR-26-165	LR-26-216	LR-26-269	
LR-26-54	LR-26-116	LR-26-169	LR-26-219	LR-26-273	
LR-26-55	LR-26-117	LR-26-170	LR-26-220	LR-26-274	
LR-26-56	LR-26-118	LR-26-171	LR-26-223	LR-26-275	
LR-26-57	LR-26-119	LR-26-172	LR-26-224	LR-26-276	
LR-26-62	LR-26-121	LR-26-173	LR-26-227	LR-26-280	
LR-26-63	LR-26-122	LR-26-175	LR-26-228	LR-26-281	
LR-26-64	LR-26-123	LR-26-176	LR-26-232	LR-26-282	

Appendix D

Table D List of SNP markers with map interval (cM) and position (cM) on the LR-26 linkage map.

LG	Marker Name	Interval	Position	LG	Marker Name	Interval	Position	LG	Marker Name	Interval	Position
1	1C224684308	0	0	1	1C303815402	0.62	30.34	1	7C157373608	0.34	81.68
1	1C292648152	0.3	0.3	1	1C304521550	0.67	31.01	1	5C223307563	0.35	82.03
1	1C319321598	2.55	2.85	1	1C301592987	1.01	32.02	1	5C232100559	0.37	82.4
1	1C217036986	1.87	4.72	1	1C117942136	0.3	32.32	1	5C244378122	0.4	82.8
1	1C217089196	0.61	5.33	1	1C301763149	0.31	32.63	1	5C244378033	0.38	83.18
1	1C317612102	1.23	6.56	1	1C299824571	0.3	32.93	1	1C88883934	0.35	83.53
1	1C317612128	0.3	6.86	1	1C300755585	0.3	33.23	1	1C19060454	0.79	84.32
1	1C317785012	0.92	7.78	1	1C299778595	0.31	33.54	1	1C181909224	1.25	85.57
1	1C317784994	0.61	8.39	1	Ctg23988110553	0.92	34.46	1	1C177752290	1.42	86.99
1	1C317453620	1.36	9.75	1	Ctg23988111526	0.64	35.1	1	5C127982752	8.26	95.25
1	1C316506809	0.7	10.45	1	1C298580243	1.63	36.73	1	5C127982816	0.7	95.95
1	1C316102216	0.31	10.76	1	1C281628693	1	37.73	1	Ctg31453220642	4	99.95
1	1C315854112	0.92	11.68	1	1C296726112	0.33	38.06	1	Ctg31453220688	0.34	100.29
1	1C315766024	0.61	12.29	1	1C295867276	0.64	38.7	1	5C216199589	2.08	102.37
1	1C313787839	0.92	13.21	1	1C292879465	0.98	39.68	1	5C216199546	0.3	102.67
1	1C313581184	0.6	13.81	1	1C292879486	0.62	40.3	1	1C110419028	0.31	102.98
1	1C329974702	0.61	14.42	1	1C182524956	0.95	41.25	1	1C174896980	0.41	103.39
1	1C312781886	2.38	16.8	1	1C287559040	1.58	42.83	1	1C205602040	1.33	104.72
1	1C311830011	0.33	17.13	1	1C238579268	1.24	44.07	1	1C238036046	2.19	106.91
1	1C309621053	0.93	18.06	1	1C277954409	0.3	44.37	1	1C238036062	0.39	107.3
1	1C310029684	0.68	18.74	1	1C233525327	0.62	44.99	1	1C211922788	1.6	108.9
1	1C310029629	0.34	19.08	1	1C278804053	0.3	45.29	1	1C203271439	1.23	110.13
1	1C309808340	0.77	19.85	1	1C278783582	0.35	45.64	1	1C220517187	0.4	110.53
1	1C308441529	0.35	20.2	1	1C278643559	0.7	46.34	1	1C226533432	1.81	112.34
1	1C308795542	0.3	20.5	1	1C290051333	0.31	46.65	1	1C5949	0.69	113.03
1	1C309440346	0.6	21.1	1	5C25194872	4.21	50.86	1	1C230680874	0.36	113.39
1	1C308055934	1.23	22.33	1	Ctg38541321468	5.06	55.92	1	1C324969811	0.35	113.74
1	Ctg22625131022	0.3	22.63	1	4C28342959	4.46	60.38	1	1C324969808	0.32	114.06
1	1C308064551	0.61	23.24	1	4C28342957	0.32	60.7	1	1C229180035	0.32	114.38
1	1C306461962	0.6	23.84	1	7C60026340	4.2	64.9	1	1C229180070	0.32	114.7
1	1C306445261	0.61	24.45	1	7C60026345	0.33	65.23	1	1C7136613	0.33	115.03
1	1C308312520	1.24	25.69	1	5C25194882	0.37	65.6	1	1C184253108	0.32	115.35
1	1C308310960	0.31	26	1	3C163221721	0.35	65.95	1	1C237646496	0.65	116
1	1C89861492	1.59	27.59	1	2C294715773	0.37	66.32	1	1C237727570	0.31	116.31
1	1C89545215	0.31	27.9	1	3C121049883	0.77	67.09	1	1C237210320	1.97	118.28
1	1C337229844	0.61	28.51	1	5C6148865	2.66	69.75	1	1C237210415	0.31	118.59
1	1C327607452	0.61	29.12	1	7C75773633	5.83	75.58	1	1C245272857	0.64	119.23
1	1C336889191	0.3	29.42	1	Ctg36668213565	4.31	79.89	1	1C159033248	1.29	120.52
1	1C303674084	0.3	29.72	1	5C106120366	1.45	81.34	1	1C248537010	0.31	120.83

Continued

LG	Marker Name	Interval	Position	LG	Marker Name	Interval	LG	LG	Marker Name	Interval	Position
1	1C248537207	0.31	121.14	1	7C238387211	1.9	163.76	1	7C215167931	0.31	192.01
1	1C252896594	0.96	122.1	1	7C237945562	2.9	166.66	1	7C195404297	0.63	192.64
1	1C248314087	0.32	122.42	1	7C237641171	0.31	166.97	1	7C214498647	0.31	192.95
1	1C250515316	1.29	123.71	1	7C237825468	0.3	167.27	1	7C212816743	1.26	194.21
1	1C256912652	0.62	124.33	1	7C237641199	0.61	167.88	1	Ctg2890009637	0.32	194.53
1	1C257392131	0.61	124.94	1	7C236100073	2.22	170.1	1	7C212032484	0.31	194.84
1	1C172759289	0.62	125.56	1	7C235942465	0.92	171.02	1	Ctg30019919274	0.32	195.16
1	1C172759472	0.32	125.88	1	Ctg1077808637	0.31	171.33	1	7C206935480	0.65	195.81
1	1C329586583	1.96	127.84	1	7C232738521	0.62	171.95	1	7C186273052	0.32	196.13
1	1C259027636	0.31	128.15	1	7C130850422	1.57	173.52	1	7C210660680	0.32	196.45
1	1C266510052	1.57	129.72	1	7C230805374	0.61	174.13	1	7C157954717	0.32	196.77
1	1C262896206	0.31	130.03	1	7C230870557	0.3	174.43	1	7C211116697	0.32	197.09
1	1C262423003	0.32	130.35	1	7C229742978	0.3	174.73	1	Ctg1280858953	1	198.09
1	1C334552586	0.32	130.67	1	7C229743010	0.61	175.34	1	7C203866645	0	198.09
1	1C68682211	0.68	131.35	1	2C17820	0.3	175.64	1	Ctg9627310008	0.7	198.79
1	1C68155076	0.32	131.67	1	7C229843752	0.31	175.95	1	Ctg1131758837	0.36	199.15
1	1C173466605	0.32	131.99	1	7C229884142	0.3	176.25	1	Ctg388659277	1.08	200.23
1	1C268319691	0.93	132.92	1	7C230632537	1.55	177.8	1	7C166948281	0.98	201.21
1	1C269167328	1.57	134.49	1	7C230632455	0.31	178.11	1	7C80563902	0.37	201.58
1	1C44870525	0.32	134.81	1	7C219476277	0.61	178.72	1	Ctg37605021147	1.33	202.91
1	1C45068478	1.3	136.11	1	7C227390157	0.31	179.03	1	7C82419400	2.14	205.05
1	1C84609203	0.62	136.73	1	7C135128831	0.31	179.34	1	7C81568015	0.35	205.4
1	1C330645066	0.63	137.36	1	7C135128747	0.3	179.64	1	7C82419175	0.35	205.75
1	7C245743874	5.58	142.94	1	7C225123287	1.55	181.19	1	7C112730708	0.7	206.45
1	7C245941983	0.91	143.85	1	7C225428266	1.24	182.43	1	7C76498894	0.38	206.83
1	7C246146392	0.3	144.15	1	7C225652447	0.61	183.04	1	7C107409608	0.75	207.58
1	7C245743756	0.92	145.07	1	7C224135765	1.24	184.28	1	7C80036021	0.35	207.93
1	7C242958570	4.23	149.3	1	7C3200675	0.31	184.59	1	7C83005571	1.43	209.36
1	7C243202125	2.52	151.82	1	7C222624888	0.31	184.9	1	7C166622280	1.06	210.42
1	7C243202065	0.3	152.12	1	7C222624897	0.3	185.2	1	7C62061089	1.36	211.78
1	7C243540350	0.6	152.72	1	7C56587895	0.3	185.5	1	7C69801663	1.33	213.11
1	7C243505003	0.31	153.03	1	7C56589081	0.31	185.81	1	7C69801756	0.32	213.43
1	7C243635439	1.22	154.25	1	Ctg78612318123	0.93	186.74	1	7C68233755	1.67	215.1
1	7C243444434	0.3	154.55	1	7C65255113	0.93	187.67	1	7C68677326	1.34	216.44
1	7C243286764	1.24	155.79	1	7C65254962	0.31	187.98	1	7C68677372	0.32	216.76
1	7C241824162	1.28	157.07	1	7C219849913	0.92	188.9	1	7C68677126	0.65	217.41
1	7C241075985	1.28	158.35	1	7C219384366	1.24	190.14	1	7C68179386	0.98	218.39
1	7C60979882	0.3	158.65	1	7C87461702	0.62	190.76	1	7C68210433	0.31	218.7
1	7C239213589	3.21	161.86	1	7C141942989	0.94	191.7	1	7C68040195	0.31	219.01

Continued

LG	Marker Name	Interval	Position
1	7C169395729	5.46	224.47
1	Ctg170662110149	2.97	227.44
1	7C63933435	0.98	228.42
1	7C231400742	0.65	229.07
1	7C63612091	0.62	229.69
1	7C62850985	0.62	230.31
1	7C62787017	0.62	230.93
1	7C62787140	0.31	231.24
1	7C63237816	1.58	232.82
1	7C62327204	1.7	234.52
1	7C62204084	0.34	234.86
1	7C62279169	0.3	235.16
1	7C61526476	1.25	236.41
1	7C61726983	0.31	236.72
1	7C61225576	2.28	239
1	7C59559866	3.06	242.06
1	7C59569773	0.35	242.41
1	7C214045805	1.04	243.45
1	7C59811512	1.69	245.14
1	7C28567985	0.66	245.8
1	7C28634348	0.3	246.1
1	7C28747122	0.62	246.72
1	7C59247180	2.58	249.3
1	7C57736059	0.94	250.24
1	7C57299123	1.03	251.27
1	7C57470433	0.34	251.61
1	7C57473692	0.3	251.91
1	7C57479258	0.31	252.22

LG	Marker Name	Interval	Position
2	1C72180116	0	0
2	1C72180102	0.76	0.76
2	1C72060026	0.74	1.5
2	1C71641192	1.49	2.99
2	1C71509799	1.05	4.04
2	1C71641139	0.35	4.39
2	1C71096778	1.95	6.34
2	1C71076197	0.75	7.09
2	1C71348823	1.01	8.1
2	1C71326284	0.34	8.44
2	1C70826417	1.35	9.79
2	1C70755583	0.32	10.11
2	1C70711069	1.27	11.38
2	1C70123998	5.81	17.19
2	1C70124054	0.31	17.5
2	1C70051275	2.98	20.48
2	1C69966912	1.58	22.06
2	1C69929415	0.33	22.39
2	1C69929561	0.34	22.73

LG	Marker Name	Interval	Position
3	2C858089	0	0
3	2C80372380	0.62	0.62
3	2C736223	0.31	0.93
3	2C736103	0.3	1.23
3	2C908673	0.93	2.16
3	2C1478031	0.62	2.78
3	2C9283274	0.3	3.08
3	2C1503409	0.61	3.69
3	2C2008730	0.61	4.3
3	2C315847279	0.92	5.22
3	2C2736442	1.77	6.99
3	2C2736287	0.7	7.69
3	2C2673992	0.34	8.03
3	2C4560483	0.95	8.98
3	2C4756977	0.92	9.9
3	2C4756961	0.31	10.21
3	2C7129376	3.98	14.19
3	2C7350294	1.26	15.45
3	2C5363139	0.61	16.06
3	2C5501567	0.3	16.36
3	2C7629038	0.61	16.97
3	2C7629217	0.61	17.58
3	2C264188544	0.61	18.19
3	2C264178520	0.92	19.11
3	Ctg10258015874	0.93	20.04
3	Ctg10258015713	0.31	20.35
3	2C9022953	0.72	21.07
3	2C9668119	1.8	22.87
3	2C10477349	0.31	23.18
3	2C7576367	0.31	23.49
3	2C11829281	1.92	25.41
3	2C11829295	0.3	25.71
3	2C13072915	2.28	27.99
3	2C13909886	0.35	28.34
3	2C13910193	0.35	28.69
3	2C207603658	0.99	29.68
3	2C14177863	1.68	31.36
3	2C124929757	2.37	33.73
3	2C108324883	0.32	34.05
3	2C15359162	0.66	34.71
3	2C16378159	1.32	36.03
3	2C17151714	0.33	36.36
3	2C17192733	0.34	36.7
3	2C17607609	1.01	37.71
3	2C18435990	0.98	38.69
3	2C19294879	0.97	39.66
3	2C20428997	1.96	41.62
3	2C20975008	0.32	41.94
3	2C19991201	1.96	43.9
3	2C173569031	1.28	45.18
3	2C107035865	1.6	46.78
3	2C173345780	0.31	47.09
3	2C23701720	0.62	47.71
3	2C254150361	0.31	48.02
3	2C25061870	1.57	49.59
3	2C24723602	0.3	49.89
3	2C31066002	0.3	50.19
3	2C29513156	1.23	51.42
3	2C29451811	0.6	52.02
3	2C29807392	1.55	53.57
3	2C26414682	0.91	54.48
3	2C26415008	0.61	55.09
3	2C72679354	1.54	56.63
3	2C32560104	0.61	57.24
3	2C38760675	0.3	57.54
3	2C85920872	0.61	58.15
3	2C240534568	0.62	58.77
3	2C37044733	0.6	59.37

Continued

LG	Marker Name	Interval	Position
3	2C192103281	0.96	60.33
3	2C127089950	0.31	60.64
3	2C49354077	0.62	61.26
3	2C51543602	0.61	61.87
3	2C74316135	0.31	62.18
3	2C58486928	0.3	62.48
3	2C158727967	0.31	62.79
3	2C294715734	0.3	63.09
3	2C57804945	0.31	63.4
3	2C92845181	2.2	65.6
3	1C338362378	0.92	66.52
3	2C296965775	0.3	66.82
3	2C145817611	0.31	67.13
3	2C188452560	0.31	67.44
3	2C188452547	0.3	67.74
3	2C55599476	0.61	68.35
3	2C179015759	1.23	69.58
3	2C153592825	1.83	71.41
3	2C162412331	0.36	71.77
3	2C162412339	0.3	72.07
3	2C153592888	0.31	72.38
3	2C162340094	0.61	72.99
3	2C170436882	0.3	73.29
3	2C172697068	1.23	74.52
3	2C172406884	0.61	75.13
3	2C158876111	0.92	76.05
3	2C158876122	0.3	76.35
3	2C174102053	0.3	76.65
3	2C127954099	1.55	78.2
3	2C180257923	0.31	78.51
3	2C175084891	0.3	78.81
3	2C176655618	0.33	79.14
3	2C234771807	1.32	80.46
3	2C182473671	0.93	81.39
3	2C76897825	0.31	81.7
3	2C187126445	0.3	82
3	2C186302101	0.35	82.35
3	2C140985157	0.71	83.06
3	2C189195967	0.3	83.36
3	2C194715124	0.31	83.67
3	2C194715213	0.61	84.28
3	2C246831988	1.01	85.29
3	2C246831999	0.33	85.62
3	2C196719805	0.33	85.95
3	2C204738031	0.31	86.26
3	2C126257431	0.92	87.18
3	2C126257361	0.32	87.5
3	2C205831010	0.65	88.15
3	2C209118606	0.61	88.76
3	2C207737054	0.69	89.45
3	2C126407089	0.34	89.79
3	2C103834365	1.58	91.37
3	2C214637541	0.3	91.67
3	2C214637365	0.35	92.02
3	2C148525446	0.71	92.73
3	2C223307435	1.24	93.97
3	1C93400577	0.93	94.9
3	2C253820633	1.4	96.3
3	2C226065606	0.7	97
3	2C195806235	0.62	97.62
3	2C240841933	0.31	97.93
3	2C194096341	0.74	98.67
3	2C241999256	1.1	99.77
3	2C243083664	0.3	100.07
3	2C61471754	0.61	100.68
3	2C61471646	0.31	100.99
3	2C115113847	0.3	101.29
3	2C259546499	0.62	101.91
3	2C263846130	1.88	103.79
3	2C267166217	0.31	104.1
3	2C273777189	1.24	105.34
3	2C278115316	0.32	105.66
3	2C278730797	0.32	105.98
3	4C215490571	0.31	106.29
3	2C283568430	0.3	106.59
3	2C285312056	0.62	107.21
3	2C20746848	0.62	107.83
3	2C193929353	0.32	108.15
3	2C193929368	0.31	108.46
3	2C37188546	1.6	110.06
3	2C285118164	0.31	110.37
3	2C289003982	0.34	110.71
3	2C288742768	0.69	111.4
3	2C290173937	2.23	113.63
3	2C292284742	0.31	113.94
3	2C293447862	0.61	114.55
3	2C293915502	0.31	114.86
3	2C296444086	0.31	115.17
3	2C297377826	1.25	116.42
3	2C297377881	0.3	116.72
3	2C299753727	1.72	118.44
3	2C300816482	1.78	120.22
3	2C300816194	0.65	120.87
3	Ctg43206825699	2.91	123.78
3	2C303760827	0.94	124.72
3	2C303207522	0.31	125.03
3	Ctg2164779007	0.31	125.34
3	2C52459109	0.68	126.02
3	2C52601761	1.06	127.08
3	2C52687558	1.6	128.68
3	2C52476189	0.32	129
3	2C52489301	0.31	129.31
3	2C304873286	0.6	129.91
3	2C304873159	0.3	130.21
3	2C305259939	0.62	130.83
3	2C305545285	0.31	131.14
3	2C306023962	1.88	133.02
3	2C306171315	0.3	133.32
3	2C305819323	0.3	133.62
3	2C205137876	0.32	133.94
3	Ctg11706422116	2.08	136.02
3	2C306399212	0.66	136.68
3	Ctg31322747511	2.29	138.97
3	2C163955570	0.63	139.6
3	2C163955498	0.3	139.9
3	2C306641437	0.92	140.82
3	2C306592894	0.92	141.74

Continued

LG	Marker Name	Interval	Position
3	2C307451126	0.61	142.35
3	2C86195591	4.23	146.58
3	2C86195613	0.3	146.88
3	2C309116384	3.19	150.07
3	2C311072396	0.61	150.68
3	2C309551746	0.91	151.59
3	2C137951859	0.61	152.2
3	2C309407028	1.03	153.23
LG	Marker Name	Interval	Position
4	3C194682482	0	0
4	3C172821932	0.61	0.61
4	3C159569466	1.24	1.85
4	3C68777841	2.11	3.96
4	3C190371312	2.1	6.06
4	3C190371260	0.3	6.36
4	3C189985223	0.92	7.28
4	3C189779352	2.19	9.47
4	Ctg149384103040	1.54	11.01
4	3C189225796	0.34	11.35
4	3C189226291	0.34	11.69
4	3C188750860	0.3	11.99
4	3C188175322	0.6	12.59
4	3C29545972	0.3	12.89
4	3C188071549	0.92	13.81
4	3C188071578	0.61	14.42
4	3C187346734	0.61	15.03
4	3C74163247	0.93	15.96
4	3C184114955	1.57	17.53
4	3C184699890	0.3	17.83
4	3C184114989	0.61	18.44
4	3C184063195	0.92	19.36
4	3C148831960	2.89	22.25
4	3C182520302	0.61	22.86
4	3C163738410	3.91	26.77
4	3C180019507	0.61	27.38
4	3C180019420	0.3	27.68
4	3C180169256	1.24	28.92
4	3C178766825	1.25	30.17
4	3C178766849	0.61	30.78
4	3C178040104	1.56	32.34
4	3C178173000	0.3	32.64
4	3C177013299	0.92	33.56
4	3C177013344	0.61	34.17
4	3C175184486	3.21	37.38
4	3C175184579	0.3	37.68
4	3C158738251	0.31	37.99
4	3C174680193	0.92	38.91
4	3C174184220	1.55	40.46
4	3C174243125	0.3	40.76
4	3C171368262	0.61	41.37
4	3C198749861	0.61	41.98
4	3C141990820	0.32	42.3
4	3C45429618	0.31	42.61
4	3C170219365	4.41	47.02
4	3C170507832	0.3	47.32
4	Ctg9165156006	2.19	49.51
4	3C128216726	1.54	51.05
4	3C4941815	0.3	51.35
4	Ctg23096415532	0.31	51.66
4	1C330173990	0.63	52.29
4	3C161979010	0.35	52.64
4	3C54524341	1.07	53.71
4	3C159613409	2.54	56.25
4	3C159613539	0.3	56.55
4	3C77925434	1.23	57.78
4	3C159411876	0.61	58.39
4	3C157709322	0.32	58.71
4	3C156095599	0.31	59.02
4	3C5322708	0.63	59.65
4	3C154059557	1.27	60.92
4	3C154059547	0.31	61.23
4	3C152970293	0.61	61.84
4	3C152970035	0.31	62.15
4	3C197610352	0.3	62.45
4	3C197610421	0.31	62.76
4	3C153584641	0.62	63.38
4	3C148907736	1.25	64.63
4	3C149217104	0.62	65.25
4	3C147944213	1.56	66.81
4	3C146943987	0.62	67.43
4	3C147185256	0.31	67.74
4	3C146025709	0.63	68.37
4	3C145101746	0.62	68.99
4	3C145155342	1.58	70.57
4	3C145155414	0.3	70.87
4	3C77116843	0.92	71.79
4	3C142418441	0.61	72.4
4	3C142313915	0.61	73.01
4	3C142220998	0.92	73.93
4	3C142220949	0.3	74.23
4	3C141391527	0.61	74.84
4	3C139615969	2.87	77.71
4	3C139883195	0.31	78.02
4	3C140443931	1.01	79.03
4	3C137631159	0.34	79.37
4	3C136691671	2.28	81.65
4	3C135291936	1.28	82.93
4	3C70798243	0.61	83.54
4	3C70798222	0.31	83.85
4	3C135627616	3.27	87.12
4	3C133816332	2.24	89.36
4	3C134238287	0.62	89.98
4	3C133258058	0.31	90.29
4	3C133326111	0.32	90.61
4	3C133796384	1.31	91.92
4	3C131340783	0.3	92.22
4	3C106176521	0.64	92.86
4	3C123383552	0.32	93.18
4	3C123158993	0.61	93.79
4	3C122856373	0.32	94.11
4	3C122856407	0.31	94.42
4	3C121049852	0.62	95.04
4	3C119015830	2.96	98
4	3C173843769	0.63	98.63
4	3C173843665	0.31	98.94
4	3C173843718	0.3	99.24
4	3C119015863	0.63	99.87

Continued

LG	Marker Name	Interval	Position
4	3C147562555	1.26	101.13
4	3C117196656	0.62	101.75
4	3C117193411	0.31	102.06
4	3C117211073	0.31	102.37
4	3C113280554	1.89	104.26
4	3C116065977	1	105.26
4	3C116898286	0.32	105.58
4	3C116381894	0.31	105.89
4	3C139750286	0.3	106.19
4	3C106586568	0.61	106.8
4	3C103799043	0.62	107.42
4	3C107756946	0.3	107.72
4	3C110523269	0.64	108.36
4	3C106660036	0.65	109.01
4	3C106746288	0.3	109.31
4	3C106748756	0.31	109.62
4	3C107270062	0.34	109.96
4	3C20248386	0.66	110.62
4	3C107009133	0.31	110.93
4	3C160678391	0.31	111.24
4	3C107009299	0.62	111.86
4	3C105194762	0.62	112.48
4	3C105194773	0.31	112.79
4	3C108738214	0.62	113.41
4	Ctg198728147103	0.61	114.02
4	Ctg198728146986	0.31	114.33
4	3C110100408	1.23	115.56
4	3C112610956	0.92	116.48
4	3C109890049	0.6	117.08
4	3C102610759	0.3	117.38
4	Ctg1778498621	0.31	117.69
4	3C109890061	0.3	117.99
4	3C105456723	0.3	118.29
4	3C112011684	0.62	118.91
4	3C112011726	0.3	119.21
4	3C104551405	0.3	119.51
4	3C108612477	0.61	120.12
4	3C185003470	0.3	120.42
4	3C102754001	0.3	120.72
4	3C102754003	0.3	121.02
4	3C100859885	0.92	121.94
4	3C33855670	0.91	122.85
4	3C98881974	0.95	123.8
4	3C127179984	0.62	124.42
4	3C186830155	0.61	125.03
4	3C94900937	1.22	126.25
4	3C73838445	1.23	127.48
4	3C93147686	0.3	127.78
4	3C189390503	0.3	128.08
4	3C139058871	0.34	128.42
4	3C91453410	0.34	128.76
4	3C91453477	0.3	129.06
4	3C178300819	0.3	129.36
4	3C197673603	0.3	129.66
4	3C90165820	0.3	129.96
4	3C87847559	0.61	130.57
4	3C197920689	0.61	131.18
4	3C155175504	0.61	131.79
4	Ctg69313823181	2.88	134.67
4	Ctg3836925937	0.61	135.28
4	3C75536302	0.31	135.59
4	Ctg1298847002	0.3	135.89
4	Ctg7281835412	0.6	136.49
4	Ctg44272810072	0.31	136.8
4	Ctg4133615893	0.93	137.73
4	3C71078920	0.3	138.03
4	1C330595654	0.61	138.64
4	3C56707201	0.3	138.94
4	Ctg68749411371	0.61	139.55
4	3C51498472	0.3	139.85
4	3C51113211	0.3	140.15
4	Ctg24050954368	0.3	140.45
4	3C57696273	0.31	140.76
4	3C63029496	0.92	141.68
4	Ctg5064119609	0.91	142.59
4	3C43515158	0.3	142.89
4	3C30327831	1.55	144.44
4	3C55247718	0.6	145.04
4	3C48567459	0.61	145.65
4	Ctg26646320290	0.3	145.95
4	Ctg5766319760	0.94	146.89
4	3C17083634	1.22	148.11
4	Ctg1863032239	0.3	148.41
4	Ctg1863032287	0.31	148.72
4	Ctg836397792	0.6	149.32
4	Ctg3492963857	4.23	153.55
4	Ctg3492961464	0.61	154.16
4	Ctg169175015	0.3	154.46
4	3C88053183	0.92	155.38
4	3C11544552	0.3	155.68
4	3C9400995	1.54	157.22
4	Ctg1381362394	0.63	157.85
4	3C60277882	2.12	159.97
4	3C60277938	0.32	160.29
4	3C16222873	0.66	160.95
4	Ctg335214942	0.37	161.32
4	3C3194023	1.92	163.24
4	3C2373396	3.23	166.47
4	3C1801706	0.91	167.38
4	3C120506	0.92	168.3
5	4C31762108	0	0
5	4C1873990	0.32	0.32
5	4C2021450	2.73	3.05
5	4C1785801	0.3	3.35
5	4C97011976	4.23	7.58
5	Ctg6477538945	2.69	10.27
5	4C5021756	0.32	10.59
5	4C5010228	0.3	10.89
5	4C6888070	3.88	14.77
5	4C7464619	1.23	16
5	4C224616079	0.92	16.92
5	4C9459162	2.21	19.13
5	4C9531256	1.23	20.36
5	4C9586692	0.62	20.98
5	4C11009048	0.99	21.97
5	4C11170420	3.75	25.72

Continued

LG	Marker Name	Interval	Position	LG	Marker Name	Interval	Position	LG	Marker Name	Interval	Position
5	4C12290907	0.6	26.32	5	4C162950645	0.92	66.56	5	4C223500452	0.92	95.11
5	4C96771780	6.46	32.78	5	4C172763043	0.62	67.18	5	4C220953160	0.3	95.41
5	4C96771786	0.6	33.38	5	4C173686359	0.3	67.48	5	4C222615301	0.35	95.76
5	4C25873880	6.07	39.45	5	4C179124271	0.66	68.14	5	4C30431779	0.34	96.1
5	1C97874303	0.61	40.06	5	4C149307514	0.32	68.46	5	1C277705015	1.94	98.04
5	4C96127366	0.91	40.97	5	4C164610003	0.61	69.07	5	1C277704893	0.3	98.34
5	4C171015573	1.23	42.2	5	4C185553604	0.3	69.37	5	Ctg8957733399	1.57	99.91
5	2C310761829	0.3	42.5	5	4C189244497	1.86	71.23	5	Ctg8957730769	0.3	100.21
5	4C110144695	0.92	43.42	5	4C191823105	0.61	71.84	5	Ctg27528130690	2.53	102.74
5	4C214018153	0.32	43.74	5	4C190314745	0.3	72.14	5	Ctg27528130836	0.31	103.05
5	4C109732526	0.65	44.39	5	4C191171002	0.31	72.45	5	Ctg4298626101	0.92	103.97
5	4C100451416	0.3	44.69	5	4C191822932	0.31	72.76	5	4C223614333	0.31	104.28
5	6C90611859	0.35	45.04	5	4C120142891	0.31	73.07	5	Ctg11113172309	1.89	106.17
5	4C120922104	0.7	45.74	5	4C154491652	0.31	73.38	5	Ctg78890513265	0.3	106.47
5	4C113497808	0.6	46.34	5	4C196235917	0.61	73.99	5	Ctg66953442284	1.55	108.02
5	4C201548793	1.57	47.91	5	4C150714841	0.3	74.29	5	4C226389215	0.31	108.33
5	4C55342046	1.24	49.15	5	4C197394214	2.2	76.49	5	4C226037621	0.3	108.63
5	4C217462603	0.3	49.45	5	4C197027590	0.3	76.79	5	4C226274840	0.3	108.93
5	4C145662855	0.65	50.1	5	4C136813052	1.31	78.1	5	4C171726977	1.23	110.16
5	4C145663086	0.32	50.42	5	4C199579008	0.32	78.42	5	4C228213819	0.63	110.79
5	4C140851118	0.32	50.74	5	4C202320039	0.92	79.34	5	4C227592092	1.95	112.74
5	4C151475753	0.92	51.66	5	4C203152322	0.3	79.64	5	4C228708544	0.31	113.05
5	4C151475847	0.31	51.97	5	4C202468807	1.96	81.6	5	4C70098852	1.58	114.63
5	4C152129786	0.62	52.59	5	4C205277619	0.32	81.92	5	Ctg1338633037	0.3	114.93
5	4C150350134	1.87	54.46	5	4C69531755	1.26	83.18	5	4C229315855	0.61	115.54
5	4C155739064	0.61	55.07	5	4C208036755	0.94	84.12	5	4C3971842	0.6	116.14
5	4C162087172	1.88	56.95	5	4C208172099	0.3	84.42	5	4C231038687	0.69	116.83
5	4C161459988	0.3	57.25	5	4C220488199	0.6	85.02	5	4C4161918	1.03	117.86
5	4C18100237	0.61	57.86	5	4C211063438	0.98	86	5	4C230699687	1.23	119.09
5	4C162856706	0.92	58.78	5	4C210800382	1.99	87.99	5	Ctg2305681334	0.61	119.7
5	Ctg6456220022	0.62	59.4	5	4C217704982	0.3	88.29	5	4C232730438	0.91	120.61
5	Ctg162752367	0.31	59.71	5	4C215325811	1.24	89.53	5	4C22369931	0.3	120.91
5	4C166216608	1.34	61.05	5	4C217793137	0.31	89.84	5	4C232208109	1.88	122.79
5	4C167884099	1.69	62.74	5	4C217793135	0.31	90.15	5	Ctg8107322074	1.55	124.34
5	4C167598160	0.64	63.38	5	4C222840157	0.63	90.78	5	Ctg2901655540	1.87	126.21
5	5C105826372	0.98	64.36	5	Ctg18189915080	0.32	91.1	5	Ctg10407934964	1.22	127.43
5	4C170358507	0.34	64.7	5	Ctg78434139264	1.23	92.33	5	Ctg20672118160	1.23	128.66
5	4C169190631	0.34	65.04	5	4C218555874	0.3	92.63	5	Ctg20672118636	0.3	128.96
5	4C171372962	0.6	65.64	5	4C131564140	1.56	94.19	5	4C235251486	0.3	129.26

Continued

LG	Marker Name	Interval	Position
5	4C234915765	0.31	129.57
5	Ctg46858738786	0.6	130.17
5	4C236293680	0.3	130.47
5	Ctg1135476151	0.61	131.08
5	4C237200460	0.3	131.38
5	4C237725310	2.22	133.6
5	Ctg20789950726	0.61	134.21
5	Ctg2988443498	0.92	135.13
5	4C238142378	0.64	135.77
5	4C238064598	0.32	136.09
5	4C8031174	0.62	136.71
5	4C238723484	0.6	137.31
5	4C238600003	0.61	137.92
5	Ctg1003533543	0.62	138.54
5	Ctg16720167182	0.72	139.26
5	Ctg9786524183	0.71	139.97
5	Ctg1523732472	2.21	142.18
5	Ctg268234508	2.55	144.73
5	Ctg5565147776	4.29	149.02
5	4C240584946	1.69	150.71
5	4C240375724	2.03	152.74
5	Ctg195367133030	0.3	153.04
5	Ctg13904451544	1.54	154.58
5	Ctg13904451517	0.3	154.88
5	4C241226435	0.31	155.19
5	Ctg2116140112	0.61	155.8
5	Ctg148623125737	0.91	156.71
5	Ctg1451768045	0.63	157.34
5	Ctg40941713870	0.31	157.65
5	Ctg40941713941	0.3	157.95
5	4C241433908	0.61	158.56
5	Ctg10372117519	2.19	160.75
5	Ctg25055226206	0.3	161.05
6	5C5610113	0	0
6	5C5840668	1.55	1.55
6	5C6148834	0.62	2.17
6	5C14128484	0.3	2.47
6	5C6720913	2.23	4.7
6	5C7041145	0.62	5.32
6	5C7041160	0.3	5.62
6	5C7041220	0.61	6.23
6	5C7158749	0.92	7.15
6	5C8234869	0.3	7.45
6	5C8180771	0.3	7.75
6	5C7822001	0.6	8.35
6	5C7358505	0.92	9.27
6	5C7358551	0.61	9.88
6	5C82228052	0.92	10.8
6	5C8641559	1.87	12.67
6	5C8865813	0.3	12.97
6	5C25156197	0.62	13.59
6	5C25069438	1.25	14.84
6	5C9316495	0.3	15.14
6	5C10136136	1.95	17.09
6	5C10113995	0.64	17.73
6	5C10695907	0.63	18.36
6	5C10993502	1.6	19.96
6	5C11229815	1.89	21.85
6	5C11297862	0.3	22.15
6	5C11902596	0.94	23.09
6	5C12702888	0.92	24.01
6	5C12119725	1.25	25.26
6	5C12245315	0.62	25.88
6	5C14495896	1.99	27.87
6	5C15929968	0.67	28.54
6	5C16378277	0.68	29.22
6	5C16957185	0.34	29.56
6	5C47151692	0.7	30.26
6	5C17759491	3.15	33.41
6	5C17334323	0.38	33.79
6	5C18148436	0.71	34.5
6	5C19027936	1.14	35.64
6	5C19590186	0.74	36.38
6	5C20435361	0.38	36.76
6	5C20884401	0.38	37.14
6	5C21082707	0.79	37.93
6	5C21998054	0.8	38.73
6	5C23542862	2.02	40.75
6	5C21482465	0.39	41.14
6	5C175966589	1.65	42.79
6	5C25194793	0.41	43.2
6	5C77592531	1.24	44.44
6	5C218219088	0.8	45.24
6	5C121396713	0.41	45.65
6	5C120409231	1.65	47.3
6	5C30183928	0.4	47.7
6	5C227690778	16.91	64.61
6	5C239603368	0.79	65.4
6	5C237957212	1.89	67.29
6	5C237952630	0.37	67.66
6	5C245977402	1.1	68.76
6	5C245977461	0.34	69.1
6	5C251631540	2.91	72.01
6	5C252339715	2.59	74.6
6	5C253204211	0.96	75.56
6	5C255894981	4.04	79.6
6	5C255895053	0.3	79.9
6	5C255708247	0.94	80.84
6	5C257989020	10.14	90.98
6	5C257785165	0.3	91.28
6	5C258022556	1.9	93.18
6	5C258022624	0.31	93.49
6	5C82241721	1.26	94.75
6	5C82241895	0.3	95.05
6	Ctg191019105942	0.93	95.98
6	5C258508373	2.87	98.85
6	5C116821597	0.6	99.45
6	5C203006198	1.05	100.5
6	5C259785418	1.05	101.55
6	5C259785424	0.61	102.16
7	6C208015017	0	0
7	6C207422080	0.31	0.31
7	6C208015224	0.31	0.62
7	6C107983854	0.33	0.95
7	6C107983845	0.33	1.28

Continued

LG	Marker Name	Interval	Position	LG	Marker Name	Interval	Position	LG	Marker Name	Interval	Position
7	6C208015104	0.34	1.62	7	6C188620963	1.23	35.97	7	6C173881930	0.91	73.01
7	6C207748880	0.61	2.23	7	6C144827500	0.34	36.31	7	6C125024685	0.3	73.31
7	6C206988181	0.92	3.15	7	6C189160325	0.33	36.64	7	6C170967670	1.07	74.38
7	6C205622518	1.24	4.39	7	6C188214822	1.88	38.52	7	6C139225753	2.55	76.93
7	6C206933248	0.62	5.01	7	6C19995759	0.95	39.47	7	6C168317974	1.55	78.48
7	6C205774634	0.63	5.64	7	6C19995962	0.32	39.79	7	Ctg66673821972	2.52	81
7	Ctg46105764560	2.57	8.21	7	6C140665456	0.31	40.1	7	6C208927955	1.87	82.87
7	Ctg46105764507	0.3	8.51	7	6C114251594	1.25	41.35	7	Ctg46841110843	1.91	84.78
7	6C206638918	0.62	9.13	7	6C114251166	0.31	41.66	7	Ctg46841110935	0.31	85.09
7	6C206635732	0.3	9.43	7	6C114321865	1.27	42.93	7	6C165344298	1.26	86.35
7	6C128160913	0.31	9.74	7	6C184518112	1.9	44.83	7	Ctg515537818	1.24	87.59
7	6C205054667	0.94	10.68	7	6C183864960	0.68	45.51	7	6C166437460	0.92	88.51
7	6C128592153	0.94	11.62	7	6C173056847	2.1	47.61	7	Ctg8540824220	0.61	89.12
7	6C128592129	0.31	11.93	7	6C173056790	0.3	47.91	7	Ctg8540824264	0.3	89.42
7	6C204013559	0.3	12.23	7	6C150708402	2.23	50.14	7	Ctg3524829812	1.89	91.31
7	6C204538018	0.31	12.54	7	6C166853174	0.62	50.76	7	6C158788278	0.34	91.65
7	6C204537992	0.31	12.85	7	6C180863476	0.31	51.07	7	6C210260241	1.76	93.41
7	6C203182179	0.31	13.16	7	6C180596996	0.92	51.99	7	6C161100440	0.32	93.73
7	6C202167267	0.63	13.79	7	6C119374686	1.56	53.55	7	6C161100442	0.31	94.04
7	6C202599804	0.34	14.13	7	6C113786566	0.93	54.48	7	6C147723668	0.32	94.36
7	6C202600041	0.32	14.45	7	6C178137833	1.03	55.51	7	Ctg13295128412	0.31	94.67
7	6C202195034	0.32	14.77	7	Ctg2479474525	0.33	55.84	7	6C190354799	1.9	96.57
7	6C132857941	0.61	15.38	7	Ctg90867711813	0.92	56.76	7	6C156453021	1.24	97.81
7	6C201507776	0.95	16.33	7	6C51495752	1.55	58.31	7	Ctg23583256706	0.31	98.12
7	6C200234677	1.59	17.92	7	6C51495733	0.31	58.62	7	6C155091517	0.61	98.73
7	6C109482770	0.92	18.84	7	6C51530645	1.55	60.17	7	6C152137108	1.24	99.97
7	6C109482862	0.61	19.45	7	6C51530698	0.3	60.47	7	Ctg18405017950	0.33	100.3
7	6C200168995	0.91	20.36	7	6C127678400	2.54	63.01	7	6C151504507	0.67	100.97
7	6C200110591	2.19	22.55	7	6C124677591	0.92	63.93	7	Ctg18805631879	0.94	101.91
7	6C198524339	3.23	25.78	7	6C17074906	0.62	64.55	7	6C116008201	0.61	102.52
7	6C198797468	0.31	26.09	7	6C177205847	1.55	66.1	7	6C139112446	0.93	103.45
7	6C196770108	2.29	28.38	7	6C175947190	0.91	67.01	7	6C145292002	0.31	103.76
7	6C196500846	0.95	29.33	7	6C35453827	0.61	67.62	7	6C145285409	0.31	104.07
7	6C194558121	0.95	30.28	7	6C125084257	0.63	68.25	7	6C142801922	0.62	104.69
7	6C193426704	0.31	30.59	7	6C35453870	0.32	68.57	7	Ctg4262318184	0.93	105.62
7	6C193189693	2.92	33.51	7	6C119014587	1.86	70.43	7	6C136910364	0.62	106.24
7	6C193189738	0.31	33.82	7	Ctg1226595825	0.92	71.35	7	6C136957394	0.62	106.86
7	6C192136324	0.31	34.13	7	6C172806661	0.37	71.72	7	6C133676781	0.62	107.48
7	6C192382242	0.61	34.74	7	6C173442516	0.38	72.1	7	6C121055278	0.31	107.79

Continued

LG	Marker Name	Interval	Position
7	6C122589657	0.31	108.1
7	6C54536669	0.3	108.4
7	6C110897612	0.64	109.04
7	6C109830594	0.31	109.35
7	6C112110542	0.62	109.97
7	6C91435756	0.3	110.27
7	6C74455035	0.64	110.91
7	6C73742262	0.31	111.22
7	6C48931486	0.31	111.53
7	6C80813477	1.23	112.76
7	6C39183463	0.31	113.07
7	6C40647820	0.3	113.37
7	6C40477135	0.61	113.98
7	6C52217403	0.3	114.28
7	4C245767073	1.24	115.52
7	Ctg23748980671	1.45	116.97
7	6C5274921	0.71	117.68
7	6C22113563	0.92	118.6
7	6C26138343	0.3	118.9
7	6C18209722	1.57	120.47
7	6C48643464	1.26	121.73
7	7C69329750	1.93	123.66
7	6C7554261	0.61	124.27
7	6C7628024	0.3	124.57

LG: Linkage group; Interval: marker interval in centiMorgan (cM); Position: marker position on LR-26 map in cM.

Appendix E:

A list of markers making up individual bins can be found in an MS Excel file here:

http://knowpulse.usask.ca/portal/LR-26-Marker-Bins-2018Mar15

Appendix F

Table F. List of QTLs among traits on SNP linkage map of LR-26.

LG	QTL region (cM)	Peak LOD	% Explained by peak	Peak position (cM)	SNP closest to peak	Trait	Critical LOD	Allele Source
1	67.09-68.09	4.18	8.2	68.09	3C121049883	TSW-CSSF13	3.3	Ler
	78.58-79.89	3.7	12.3	78.58	Ctg36668213565	TRFO-CSSF15	3.3	Lcu
	80.89-81.34	8.4	17.7	80.89	5C106120366	DTE-CSSF13	3.4	Lcu
	80.89-81.68	13.8	10.1	80.89	5C106120366	DTE-STH15	3.3	Ler
	83.53-85.32	7.1	17.7	83.53	1C88883934	PH-CSSF13	3.1	Lcu
	83.53-85.32	5.6	10.4	84.32	1C19060454	PH-CSSF14	3.1	Lcu
	83.18-83.53	5	11.6	83.53	1C88883934	PH-CSSF15	3.2	Lcu
	94.1	3.9	6.5	94.1	5C127982752	TSW-CSSF15	3.3	Lcu
	96.79-98.8	3.5	4.7	98.8	Ctg31453220642	DTE-CSSF14	3.2	Lcu
	101.52-103.83	4.6	12.4	102.83	1C110419028	VP-CSSF14	3.1	Ler
	104.06-109.01	7.2	7.6	107.01	1C205602040	TSW-CSSF14	3.3	Lcu
	106.06	3.7	4.4	106.06	1C205602040	TSW-CSSF13	3.3	Lcu
	124.11	3.4	7.3	124.11	1C159033248	Sucrose-CSSF15	3.2	Lcu
	132.68-132.99	5.3	13.7	132.99	1C262896206	PH-STH15	3.3	Lcu
	157.21-181.06	4.6	9.7	170.83	7C237641199	PH-CSSF13	3.1	Lcu
	157.21	3.1	5.5	157.21	7C243635439	PH-CSSF14	3.1	Lcu
	199.72-200.04	5.7	15	200.04	7C211116697	Sucrose-CSSF14	3.2	Lcu
	201.04-203.74	6	10.9	202.74	Ctg9627310008	DTE-STH15	3.3	Lcu
	201.04-203.74	4.2	7.9	201.74	Ctg9627310008	DTF-CSSF14	3.2	Lcu
	201.04	4.5	7.9	201.04	7C203866645	DTF-STH15	3.3	Lcu
	201.04-203.74	5.2	13.8	201.74	Ctg9627310008	VP-CSSF13	3.1	Lcu
	201.04-203.74	8.9	16.8	201.74	Ctg9627310008	DTF-CSSF13	3.2	Lcu
	201.04-203.74	10.9	13.2	201.74	Ctg9627310008	TSW-CSSF13	3.3	Lcu
	206.49-207.57	5.1	4.7	207.49	Ctg388659277	TSW-CSSF14	3.3	Lcu
	206.49-207.57	6.5	11	207.49	Ctg388659277	TSW-CSSF15	3.3	Lcu
	206.49-207.49	8.2	12.2	207.49	Ctg1131758837	DTE-CSSF13	3.4	Lcu
	206.49-207.57	8.7	14.9	206.49	Ctg1131758837	DTE-CSSF14	3.2	Lcu
	207.49-208.92	4	3.7	207.49	Ctg388659277	TSW-STH15	3.2	Lcu
LG	QTL region (cM)	Peak LOD	% Explained by peak	Peak position (cM)	SNP closest to peak	Trait	Critical LOD	Allele Source
3	0-10.21	4.1	6.5	1.23	2C736103	PH-CSSF14	3.1	Lcu
	63.07-64.4	3.5	8.2	63.4	2C57804945	Sucrose-CSSF14	3.2	Lcu
	64.4-67.74	3.8	5.1	66.52	1C338362378	DTE-CSSF13	3.4	Lcu
	66.52	3.1	9.4	66.52	1C338362378	Sucrose-STH15	3	Lcu
	119.44-122.87	6.6	6.9	121.87	2C300816194	TSW-CSSF14	3.3	Lcu
	120.22-122.87	4.8	13.9	122.87	2C300816194	TRFO-STH15	3.3	Lcu
	121.87-125.34	3.4	3.3	123.78	Ctg43206825699	TSW-STH15	3.2	Lcu
	123.78-125.34	3.6	5.7	124.72	2C303760827	DTF-CSSF13	3.2	Lcu
	129.91-133.02	3.5	3.5	133.02	2C306023962	TSW-STH15	3.2	Lcu
LG	QTL region (cM)	Peak LOD	% Explained by peak	Peak position (cM)	SNP closest to peak	Trait	Critical LOD	Allele Source
4	7.28-10.47	5.5	10.7	8.28	3C189779352	DTF-CSSF14	3.2	Lcu
	9.28-9.47	3.7	6.1	9.47	3C189779352	DTF-CSSF13	3.2	Lcu
	14.42-15.03	3.6	8.4	14.42	3C188071578	VP-CSSF14	3.1	Lcu
	51.05-51.35	4.4	8.2	51.35	3C4941815	DTF-CSSF13	3.2	Lcu
	51.66	3.3	5.4	51.66	Ctg23096415532	PH-CSSF14	3.1	Ler
	51.05-56.55	4.9	5	55.71	3C159613409	TSW-CSSF13	3.3	Lcu
	52.64-55.71	5.6	5.2	53.64	3C54524341	TSW-CSSF14	3.3	Lcu
	61.23-61.84	4.4	4.8	61.23	3C154059547	TSW-STH15	3.2	Lcu
	98.94-109.96	4.9	11.7	107.42	3C103799043	TRFO-CSSF15	3.3	Ler
	109.31	3.5	8.6	109.31	3C106746288	Sucrose-CSSF15	3.2	Ler
	111.86-121.02	3.8	11.4	115.56	3C110100408	Sucrose-STH15	3	Ler
	140.15-141.68	3.4	10.5	140.76	3C57696273	Sucrose-STH15	3	Ler
	157.85-158.85	3.7	7.8	157.85	Ctg1381362394	Sucrose-CSSF15	3.2	Ler

Continued

LG	QTL region (cM)	Peak LOD	% Explained by peak	Peak position (cM)	SNP closest to peak	Trait	Critical LOD	Allele Source
5	56.07-58.78	4.6	10.9	57.86	4C18100237	TRFO-CSSF15	3.3	Lcu
	59.4-76.79	4.8	7	69.07	4C164610003	TSW-CSSF15	3.3	Lcu
	67.18-67.48	6.5	16.5	67.48	4C173686359	VP-STH15	3.2	Lcu
	67.48	3.5	5.4	67.48	4C173686359	DTF-STH15	3.3	Lcu
	139.26-151.71	4.8	8.1	146.73	Ctg268234508	DTF-STH15	3.3	Lcu
	144.73-151.71	3.6	5.8	150.71	4C240584946	DTF-CSSF14	3.2	Lcu
	154.04-154.58	3.9	10.4	154.58	Ctg13904451544	RP-STH15	3.1	Ler
LG	QTL region (cM)	Peak LOD	% Explained by peak	Peak position (cM)	SNP closest to peak	Trait	Critical LOD	Allele Source
6	44.2-47.3	3.8	9.8	45.24	5C218219088	TRFO-CSSF15	3.3	Lcu
	47.3-64.61	6.5	11.8	57.7	5C227690778	TSW-STH15	3.2	Lcu
	50.7-57.7	4	12.2	53.7	5C30183928	DTE-CSSF14	3.2	Lcu
LG	QTL region (cM)	Peak LOD	% Explained by peak	Peak position (cM)	SNP closest to peak	Trait	Critical LOD	Allele Source
7	29.33-30.28	4.3	12.1	30.28	6C194558121	TRFO-STH15	3.3	Lcu
	30.28	3.5	9.3	30.28	6C194558121	TRFO-CSSF15	3.3	Lcu
	37.64-45.51	4.1	8.8	39.79	6C19995962	PH-CSSF13	3.1	Lcu
	43.35-41.66	3.7	3.6	41.66	6C114251166	TSW-CSSF13	3.3	Lcu
	47.61-49.91	3.8	3.8	49.91	6C127678400	TSW-CSSF13	3.3	Lcu
	50.14-53.55	3.8	3.8	53.55	6C119374686	TSW-CSSF13	3.3	Lcu
	51.99	3.2	7.1	51.99	6C180596996	PH-CSSF13	3.1	Lcu
	55.48-56.71	4.1	4.1	55.84	Ctg2479474525	TSW-CSSF13	3.3	Lcu
	57.76-58.31	3.7	3.8	57.76	Ctg38541321468	TSW-CSSF13	3.3	Lcu
	60.17	3.1	7	60.17	6C51530645	PH-CSSF13	3.1	Lcu
	60.47-67.01	5.1	4.6	63.01	6C127678400	TSW-CSSF13	3.3	Lcu
	61.47-64.55	4.3	3.8	63.01	6C127678400	TSW-CSSF14	3.3	Lcu
	63.93	3.4	5.1	63.93	6C124677591	TSW-CSSF15	3.3	Lcu
	68.25-74.38	3.9	3.5	71.72	6C172806661	TSW-CSSF13	3.3	Lcu
	71.35-71.72	4.95	5.6	71.7	6C172806661	TSW-STH15	3.2	Lcu
	91.31-99.97	4.3	10.9	93.73	6C161100440	VP-CSSF15	3.1	Ler
	120.47-122.73	4.7	13.5	122.73	6C48643464	VP-CSSF15	3.1	Lcu
	124.27-124.57	3.6	7.9	124.57	6C7628024	VP-STH15	3.2	Ler

LG: Linkage group; cM: centiMorgan; Lcu: *Lens culinaris*; Ler: *L. ervoides*.

LG, QTL position, peak position and SNP markers all based on LR-26 linkage map.

Traits include: days to emerge (DTE) phenotyped in Crop Science Seed Farm (CSSF) in 2013 (DTE-CSSF13), CSSF in 2014 (DTE-CSSF14) and Sutherland farm (STH) in 2015 (DTE-STH15); days to flower (DTF) phenotyped in CSSF in 2013 (DTF-CSSF13), CSSF in 2014 (DTF-CSSF14) and STH in 2015 (DTF-STH15); vegetative period (VP) phenotyped in CSSF in 2013 (VP-CSSF13), CSSF in 2014 (VP-CSSF14), CSSF in 2015 (VP-CSSF2015) and STH in 2015 (VP-STH15); reproductive period (RP) phenotyped in STH in 2015 (RP-STH15); plant height (PH) phenotyped CSSF in 2013 (PH-CSSF13), CSSF in 2014 (PH-CSSF14), CSSF in 2015 (PH-CSSF2015) and STH in 2015 (PH-STH15); thousand seed weight (TSW) phenotyped in CSSF in 2013 (TSW-CSSF13), CSSF in 2014 (TSW-CSSF14), CSSF in 2015 (TSW-CSSF2015) and STH in 2015 (TSW-STH15); sucrose concentration (Sucrose) phenotyped in CSSF in 2014 (Sucrose-CSSF13), CSSF in 2014 (Sucrose-CSSF15) and STH in 2015 (Sucrose-STH15); and total raffinose family oligosaccharides concentration (TRFO) phenotyped in CSSF in 2015 (TRFO -CSSF15), and STH in 2015 (TRFO -STH15).